EFFECT OF RELIABILITY ON CUE EFFECTIVENESS AND DISPLAY SIGNALING

BY

JAMES LOUIS MERLO

B.S., United States Military Academy, 1989

THESIS
Submitted in partial fulfillment of the requirements
for the degree of Master of Science in Psychology
in the Graduate College of the
University of Illinois at Urbana-Champaign, 1999

19990518 078

DISTRIBUTION STATEMENT A
Approved for Public Release
Distribution Unlimited

Urbana, Illinois

DTIC QUALITY INSPECTED 4

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 074-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503

1. AGENCY USE ONLY (Leave blank)

2. REPORT DATE
7 MAY 1999

3. REPORT TYPE AND DATES COVERED

4. TITLE AND SUBTITLE
EFFECT OF RELIABILITY ON CUE EFFECTIVENESS AND DISPLAY SIGNALING

5. FUNDING NUMBERS

6. AUTHOR(S)
JAMES L. Merlo +
CHRISTOPHER D. WICKENS

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
University of Illinois, MS program @ U of I Urbana-Champaign
Army Research Lab

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release: Distribution is unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 Words)*

14. SUBJECT TERMS

15. NUMBER OF PAGES

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT

18. SECURITY CLASSIFICATION OF THIS PAGE

19. SECURITY CLASSIFICATION OF ABSTRACT

20. LIMITATION OF ABSTRACT

-AQ #

☐ _____ copies are being forwarded. Indicate whether Statement A, B, C, D, E, F, or X applies.

☒ DISTRIBUTION STATEMENT A:
APPROVED FOR PUBLIC RELEASE: DISTRIBUTION IS UNLIMITED

☐ DISTRIBUTION STATEMENT B:
DISTRIBUTION AUTHORIZED TO U.S. GOVERNMENT AGENCIES ONLY; (Indicate Reason and Date). OTHER REQUESTS FOR THIS DOCUMENT SHALL BE REFERRED TO (Indicate Controlling DoD Office).

☐ DISTRIBUTION STATEMENT C:
DISTRIBUTION AUTHORIZED TO U.S. GOVERNMENT AGENCIES AND THEIR CONTRACTORS; (Indicate Reason and Date). OTHER REQUESTS FOR THIS DOCUMENT SHALL BE REFERRED TO (Indicate Controlling DoD Office).

☐ DISTRIBUTION STATEMENT D:
DISTRIBUTION AUTHORIZED TO DoD AND U.S. DoD CONTRACTORS ONLY; (Indicate Reason and Date). OTHER REQUESTS SHALL BE REFERRED TO (Indicate Controlling DoD Office).

☐ DISTRIBUTION STATEMENT E:
DISTRIBUTION AUTHORIZED TO DoD COMPONENTS ONLY; (Indicate Reason and Date). OTHER REQUESTS SHALL BE REFERRED TO (Indicate Controlling DoD Office).

☐ DISTRIBUTION STATEMENT F:
FURTHER DISSEMINATION ONLY AS DIRECTED BY (Indicate Controlling DoD Office and Date) or HIGHER DoD AUTHORITY.

☐ DISTRIBUTION STATEMENT X:
DISTRIBUTION AUTHORIZED TO U.S. GOVERNMENT AGENCIES AND PRIVATE INDIVIDUALS OR ENTERPRISES ELIGIBLE TO OBTAIN EXPORT-CONTROLLED TECHNICAL DATA IN ACCORDANCE WITH DoD DIRECTIVE 5230.25, WITHHOLDING OF UNCLASSIFIED TECHNICAL DATA FROM PUBLIC DISCLOSURE, 6 Nov 1984 (Indicate date of determination). CONTROLLING DoD OFFICE IS (Indicate Controlling DoD Office).

☐ This document was previously forwarded to DTIC on _____ (date) and the AD number is _____.

☐ In accordance with provisions of DoD instructions, the document requested is not supplied because:

☐ It will be published at a later date. (Enter approximate date, if known).

☐ Other. (Give Reason)

DoD Directive 5230.24, "Distribution Statements on Technical Documents," 18 Mar 87, contains seven distribution statements, as described briefly above. Technical Documents must be assigned distribution statements.

James L. Merlo
Print or Type Name

7 MAY 99
Authorized Signature/Date

(217) 384-0599
Telephone Number

UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

THE GRADUATE COLLEGE

APRIL 1999
(date)

WE HEREBY RECOMMEND THAT THE THESIS BY

JAMES LOUIS MERLO

ENTITLED EFFECT OF RELIABILITY ON CUE EFFECTIVENESS AND DISPLAY SIGNALING

BE ACCEPTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR

THE DEGREE OF MASTER OF SCIENCE

Director of Thesis Research

Head of Department

Committee on Final Examination†

Chairperson

† Required for doctor's degree but not for master's.

O-517

EFFECT OF RELIABILITY ON CUE EFFECTIVENESS AND DISPLAY SIGNALING

James Louis Merlo, M.S.
Department of Psychology
University of Illinois at Urbana-Champaign, 1999
Christopher D. Wickens, Advisor

Twenty Army personnel using either a hand-held display (HHD) or a helmet-mounted display (HMD) were asked to detect, identify, and give heading information for targets hidden in the far domain while performing a monitoring task in the near domain. Both displays had target cueing present for half of the trials, with the precision of the target cues varied across blocks. While all of the experimental blocks contained some trials with attentional cueing that was extremely accurate, placing the transparent cueing symbology on top of the target, some of the blocks had trials that also contained cues with degraded precision cueing, the cueing symbology located up to 22.5 degrees from the target center, or poor precision cueing, the cueing symbology located up to 45 degrees in visual angle from the target center. Explicit display of the precision reliability of the cues was attempted in order to help subjects diffuse attention during trials that lacked extremely accurate precision in cueing. During the last experimental block the automated target cueing catastrophically fails, causing users to experience costs with overtrust in automation, and then costs associated with undertrust in the automation for subsequent trials. Analysis of the results were conducted looking at two specific costs of attentional cueing that we define as type A (attention) costs and type T (trust) costs. Results show that spatially accurate cued targets were found faster than uncued targets. Targets were always found fastest on average under the cued condition while using an HMD, however, when targets are uncued the salience of the

different target types caused differences in the detection rate between the display platforms suggesting a scan/clutter tradeoff between HMDs and HHDs. Attention cueing induced a type A (attention) cost, shown by the low detection rate of a higher priority but uncued target when it was simultaneously presented with a lower priority cued target. For low salience targets, cueing that lacked spatial precision by more than 22.5 degrees of visual angle from the target center did not promote faster detection than if the target was uncued, however, the imprecise cueing remained beneficial for detection of the high salience targets even at the poor precision level. Type A and T costs were observed when the automation failed unexpectedly, with subjects initially showing signs of overtrust of the cueing information, and then on subsequent trials tending to undertrust the cueing information, with restored trust seemingly returning after a few reliable trials. Lastly, failures in automation seemed to mediate the effects of type A costs, as the detection rate of the higher priority but uncued targets increased.

Acknowledgments

The author would like to thank Ron Carbonari for his time, expertise and patience in programming the simulation. Many thanks to Drs. Christopher Wickens and Art Kramer for their suggestions and theoretical contributions and a heartfelt thanks to Michelle Yeh, Heather Pringle, and Jake Seagull for their countless hours of teaching, coaching, and mentoring. Thanks also to the great scientists at NCSA who helped us keep the computers running, especially Kaz, Jim, and Hank. Finally, I would like to personally thank all of the officers, non-commissioned officers and cadets who participated in the experiment for giving up their precious time to contribute to our research. Research funding for the project was provided by a grant from the Army Research Lab.

Table of Contents

List of Figures vii
1. Introduction 1
2. Helmet Mounted Displays (HMDs) 2
3. Visual Attention 6
   3.1 Object based and Space based theories of attention 6
   3.2 Orienting Attention with Cueing 8
   3.3 Attentional tunneling 11
4. Cue Reliability 13
   4.1 Costs of less than reliable cueing 13
   4.2 Human Trust in Automation 16
   4.3 Implicit variables 21
   4.4 Explicit variables 23
   4.5 Defining Reliability/Uncertainty 27
5. Summary 30
6. Method 33
   6.1 Subjects 33
   6.2 The Task 33
   6.3 Apparatus and Materials 34
   6.4 Experimental Design 38
   6.5 Procedure 40
   6.6 Performance Measures 43
7. Results 45
   7.1 The effects of spatially accurate, reliable cueing presentation 46
   7.2 The effects of spatial precision on reliable cueing 50
   7.3 Attention to expected vs. unexpected events (type A costs) 51
   7.4 Head movement during target search 53
   7.5 The effects of cueing on azimuth judgment 56
   7.6 The effects of cueing on secondary task performance 57
   7.7 Trust in automation during and after failures (type T costs): Block 11 data analysis 58
8. Discussion 64
References 71

List of Figures

Figure 2.1 Helmet Mounted Displays (HMDs) 3
Figure 3.3.1 Illustration of a type A (attentional) costs 11
Figure 4.1.1 Illustration of a type T (trust) costs 15
Figure 4.2.1 Reliability calibration 17
Figure 6.2.1 Targets 34
Figure 6.3.1 Displays - (a) HMD example. (b) HHD example 36
Figure 6.5.1 Differences in precision reliability 40
Figure 6.5.2 The upper portion of the wand 41
Figure 6.5.3 The different levels of cue imprecision 43
Figure 7.1.1 Detection times for both landmines and soldiers when target is cued or uncued 46
Figure 7.1.2 Detection times for landmines and soldiers when target is accurately cued or uncued 47
Figure 7.1.3 Detection times for accurately cued tanks 48
Figure 7.2.4 Effects of cue precision 50
Figure 7.3.4 Subject accuracy for the target identification task 52
Figure 7.4.4 Effects of cueing on head movement 54
Figure 7.4.5 The effects of display and cue precision on head movement along the x and y axes 55
Figure 7.5.6 (a) The effects of display and cueing on azimuth judgments. (b) The effects of display and cue precision on azimuth judgements 56
Figure 7.6.7 The effects of display type and cueing on secondary task performance 57
Figure 7.7.8 The results of trials in block 11 where the automation contains catastrophic failures 60

1. Introduction

The "information age" has changed almost every aspect of the way we perform our jobs and conduct our everyday lives. As our society has become more automated and computerized, so too has the way that we wage our wars. These technological advances were expected and somewhat welcomed in the Army's tactical operations centers (TOCs), fire direction control centers (FDCs) and other areas where massive amounts of data, coordinates, and information are managed manually and compiled mostly by hand. However, as technology is advancing, the 21st Century Land Warrior System will require the *ground soldier* to be able to integrate multiple channels of electronic information effectively and efficiently, as well as continue to perform his historical tactical role on the battlefield (National Research Council, 1997)

The TOCs and FDCs use more traditional display platforms, in part, because the personnel who are using information displays in these tactical centers are primarily concerned with the information being presented and are not involved in extraneous tasks like navigation, self protection, etc. In contrast, the information demands of the ground soldier presents a greater challenge in that he not only has to be able to readily gather and process the information being presented, but must also perform a myriad of additional concurrent tasks. The soldier must maintain situation awareness, both globally and locally, in areas such as land navigation, target identification and location and usually must do all this while moving in terrain that can range from mountains to swamps, from scorching heat to subzero climates.

Currently two major ways to visually display electronic information to the ground soldier are being researched. One is to present the information in a more traditional display platform like a hand-held display (HHD), the other is to present the information in a helmet-mounted display (HMD).

2. Helmet Mounted Displays (HMDs)

An HMD is a display platform mounted on the user's helmet, which allows the user to view the display through an eyepiece or visor. The display can be transparent, allowing the user to see the external environment with the data from the HMD superimposed or overlaid in the users' forward field of view (FFOV), or the display can be opaque allowing the user to only see that part of the environment that is not obscured by the presentation of the HMD. An HMD with an opaque display might appear as a high resolution 15-17 inch monitor that is about two feet from the eye, with the rest of the environment visible except that which falls directly behind the area of the display.

Independent of display format, the displays can also differ in the way they are presented to the user. There are currently three different ways that HMDs might be presented to the user: monocular, biocular, and binocular:

(1) A monocular display presents the image to only one eye, with the other eye having an unaided view of the far domain.

(2) A biocular display presents the same image to both eyes so that the resulting view is a two-dimensional display.

(3) A binocular display presents a slightly offset view to each eye allowing the user to perceive the image with stereopsis depth cues.

Figure 2.1 shows two different types of monocular opaque helmet mounted displays that are currently being used or tested.

Figure 2.1 Helmet Mounted Displays (HMDs). The man on the left is wearing a monocular, opaque head mounted display (also known as a wearable computer). The model shown is an MA IV manufactured by Xybernaut. The soldier on the right is wearing a monocular opaque helmet mounted display coupled with a computer in his backpack. The solider shown on the right is participating in a program of research by Sytronics Inc. called digitally aided soldier for human engineering research (DASHER).

Many of the benefits that are being predicted for the use of the HMD by the land warrior have already been taken advantage of in the use of head-up displays (HUDs) in the aircraft cockpit. In fact, all of the tactical fighter aircraft in the United States are equipped with HUDs (Roscoe, 1987). Similar to the HMD, the HUDs for aircraft present a virtual image within the operator's forward field of view (FFOV) allowing the pilot of the aircraft the ability to view the HUD symbology presentation overlaid upon the outside environment. The information or imagery that is being presented by the HUD is collimated in an effort to project the imagery at optical infinity (i.e. at the same distance as the information in the far domain). As with the HMD, the HUD symbology might include information pertinent to the operation and the navigation (of the aircraft), which would otherwise have to be accessed "head down", looking inside the cockpit where the instrumentation is located.

There are at least two benefits that both the HUD and the HMD share. First, displaying pertinent information in the user's FFOV reduces the head-down time. By presenting information like maps or directions in the FFOV, the user is able to reduce the amount of time it might take to scan a map that is usually located in a head-down position, and concurrently avoid obstacles or monitor rough terrain which are usually only viewable in a head-up position. In routine land navigation, this scanning between map (near domain) and terrain (far domain), might occur numerous times as the comparison of map and terrain are used to determine the user's location.

Secondly, the need to reaccommodate or refocus the eyes between near domain (symbology or presentation) and the far domain (the environment) is eliminated since the two domains both appear at the same optical distance when the image is collimated. This focusing of the eyes that occurs when the user is trying to look at objects that are at different focal distances may take as long as four seconds (Larry and Elworth, 1972).

The HMD works in a similar manner as the HUD. The advantage the HMD has over the HUD is that it is less restrictive since the display is presented on a visor or eyepiece that is attached to the user's helmet rather than confined to the windscreen of the airplane cockpit. The HMD allows freedom of movement and head rotation within the environment and can actually provide performance advantages over the HUD. Osgood and Wells (1991) showed that the use of a head-tracked HMD produced superior performance relative to that of the HUD for air-to-ground target detection because the pilots could move their head more freely to search for terrain cues. When the HMD is fitted with a head-tracking device, the head tracker can work in conjunction with the display generation allowing the user's head movements to be considered in the presentation of symbology or information to the user. The specific advantage of this head-tracked/image update is realized in the presentation of world-referenced imagery.

The benefits of using world-referenced imagery, given the appropriate use of head trackers, is that information can be presented in such a manner that it has direct spatially defined referents in the outside world, thus the presentation of the imagery can become conformal, with a direct one-to-one mapping between projected imagery and the external world. Such conformal imagery could be used to show contour lines on terrain (Behringer, 1999), navigation routes or even cue known enemy target locations (Yeh, Wickens, & Seagull, 1999). Although these types of information could be presented on a head-down display, such imagery would require a greater degree of mental transformation or mental rotation to identify and orient to a direct position in space. Research with HUDs has shown considerable benefit for conformal imagery when presented heads-up versus heads-down (Wickens and Long, 1995).

While there are benefits for the HMD/HUDs there are also cost associated with presenting information in this heads up manner. We will now address some of these costs and concerns, related to the area of visual attention.

3. Visual Attention

Imagine a soldier searching for targets (focusing attention on the far domain), while also attempting to examine navigation information on a display (focusing attention on the near domain). One issue that arises is whether the superimposition of symbology (via HMD), can increase the soldier's ability to divide attention between these two channels and tasks.

Two varieties of attention that must be considered when addressing the issue of HUD and HMD technology are divided and focused attention. Researchers have shown that the display characteristics or symbology used in HUD/HMD's can influence performance on both focused and divided attention tasks. Factors like the symbology's salience and intensity, as well as spatial proximity and the degree of "objectness" created by the features in the field of view all influence attention allocation and division in processing the images viewed in an HUD/HMD (Ververs and Wickens, 1998).

3.1 Object based and Space based theories of attention

There are two different theories of visual attention that model the way that attention is allocated to different regions of space: space-based and object-based (Kramer and Jacobson 1991).

In space-based theories, attention is often compared to a spotlight and information which falls inside the spotlight's beam is more efficiently processed than that which is outside (Wickens, 1992). Eriksen and Yeh (1985) and Eriksen and St. James (1986) proposed a zoom lens model of the visual attentional field in which the spotlight of attention is of a variable size. This model suggests that when attention is directed or cued to a specific stimulus, the attentional field effectively zooms in on the stimulus, increasing resource density and decreasing field size until the

resource concentration is sufficient to make the required discrimination or extraction of detail and information. This metaphorical "spotlight" or "zoom lens" can be directed or cued through the use of different techniques within displays. Eriksen and St. James (1986) additionally reported that the size of the visual attentional focus may also be manipulated and that the area of attentional focus is characterized by an essentially even distribution of processing resources. Johnston and Dark (1986) also suggested that the area of the attention spotlight could change. As one example, Barriopedro and Botella (1998) showed that when a subject knew where in the visual field a stimulus would occur, the subject actually narrowed his or her attentional focus to that area.

In contrast to space-based theories, in object-based theories, attention to is based more on grouping factors, or objects that are processed sequentially. The physical proximity of the attributes within some object is not the primary factor driving attentional processing. Contour, color, and movement also serve to influence the distribution of attention, and these attributes within an object are processed in parallel (Kramer and Jacobson 1991). In the sequential process of attention distribution or deployment there is a preattentive process in which displayed information is separated into perceptual parts and then a second stage in which focal attention is used to analyze specific objects sequentially, in more detail.

The fundamental difference between the object-based theory and the space-based theory relates to which competing characteristics of the visual field determine how attention is directed. Object-based theories suggest that it is the target itself that separates the visual field into distinguishable components, whereas the space-based theory suggests that it is the target's spatial location within the visual field to which attention is allocated.

Now suppose that there is a cue in the HMD that points to a likely target in the far domain, thus requiring divided attention between the near domain cue and the fair domain target. Does a HMD help here?

3.2 Orienting Attention with Cueing

How might cueing affect the distribution of attention adopting a space-based model? Posner, Snyder and Davidson (1980) distinguished between two different aspects of the attentional system: *orienting* and *detecting*. Orienting is that aspect of attention that denotes where and in what direction attention is focused, a sort of selection of a position in space. Detection on the other hand is when there is contact between the attentional system and the signal. From these two different aspects of attention, Posner et al. conclude that the efficiency of target detection is directly affected by orienting and therefore, orienting must either precede detecting or occur at the same time.

There are many techniques that can direct attention within displays. The use of an attentional cue is one such method to guide attention. Attentional cues have three major properties: *physical salience*, *precision*, and *reliability*. We will address the properties of physical salience and cue precision in this section and then address the remaining property, cue reliability, in section 4.

The physical salience of an attentional cue refers to those attributes of the cue that allow it to guide or direct attention. These attributes of a cue might include the cue form, such as an arrow pointing in the direction that attention should be oriented, or the use of color coding, that makes the target object "pop-out" from the non-target items. Examples of physical salient cues

include highlighting, decluttering, and directional guidance. We will now review some of these techniques.

Studies have shown that highlighting is an effective way to guide attention (Brown, 1991; Donner, O'Brien & Rudisill, 1991; Smith and Goodwin, 1972; Van Orden and DiVita, 1993). For example, Fisher, Coury, Tengs, & Duffy (1989) showed that on average, subjects are able to find a highlighted target word among 35 other distracter words in a display almost four seconds faster than in a display with no highlighting.

While Fisher et al. (1989) used luminance intensification to highlight target words, another technique in highlighting is the use of flicker. Van Orden, DiVita, & Shim (1993) found that search times were also reduced when flicker is used to highlight color-coded symbols. Their use of luminance and flicker reduced search times in a simple search task for geometric shapes and did not increase search times for symbols that were not highlighted.

Decluttering is yet another effective way to indirectly orient attention to relevant components on a display. Drury and Clement (1978) found that search time was dependent on the number of background characters; thus, one way to reduce the pronounced effect of background characters is to remove some of these "distractors," implicitly orienting attention to the remaining characters. Hofer, Palen, and Possolo (1993) conducted a study with aviators, comparing performance on a search task using either a decluttered approach chart in which some of the landing information was removed or a standard, high information density approach chart. They reported that decluttering shortened the visual search process and that pilots performed tasks faster and with higher accuracy with the decluttered display than with the standard one.

While research has shown a benefit for cueing/orienting techniques like highlighting, research has also shown that the precision with which highlighting is implemented will influence

its effectiveness. Fisher and Tan (1989) showed that the type of highlighting, and the probability that the subjects attend first to the highlighted option, both have direct implications on whether or not there will be benefit from the highlighting. The probability that the subject attends first to the highlighted option depends to the second property of attentional cueing: *precision,* which is the ability of the cue to reduce the search space or decrease the number of search options. As part of their experiment discussed earlier, Fisher, Coury, Tengs, & Duffy (1989) also examined the effects of varying levels of cue precision. Fisher et al. manipulated the precision of the cue on some trials by not only highlighting the target word among the distracters, but also highlighting a few of the distracter words as well. In some conditions, they highlighted as many as 12 words (11 distracters and the one target word) in their 35 word display. Subjects showed an increase in detection time for the target word in the condition that highlighted more than just the target word (i.e. low cue precision) when compared to the trials that only the target word was highlighted (i.e. high cue precision). However, the subjects still had faster detection times in the low cue precision condition than the condition that had no target cueing at all, because the subjects' overall search space had been reduced (i.e. the subject only searched a highlighted 12 word list as compared to the 36 word list). Thus, although a reduction in the precision of the cue increased response time for target detection, these times were still faster than those obtained when no cue was available.

In general, the preponderance of the literature shows that the use of cues with reliable attention cueing, even if imprecise, results in improved performance (as measured by reaction time) relative to no cueing at all (e.g. Fisher, Coury, Tengs, & Duffy, 1989; Fisher and Tan, 1989; Hofer, Palen, and Possolo, 1993). However, there are sometimes costs associated with attention cueing. As previously discussed, attention cueing generally involves directing attention to a location in space. A phenomenon that can occur as a result is cognitive or attentional tunneling,

which we discuss in detail in the following section. Attentional tunneling is when a user has an undesirable and extended fixation on an object, spatial position or aspect of the display. In some cases, the fixation can occur on HUD symbology, as observed in Wickens and Long (1995) or for the current study, it may be the region that is highlighted or cued.

3.3 Attentional tunneling

Attention tunneling is a misallocation of attentional resources. The cost associated with a valid cue that directs attention away from another equally or more important event or location and thereby causes an unwanted attention tunneling, will be labeled a type A (attentional) cost (Wickens, Pringle, & Merlo, 1999). A schematic example of this cost is shown in figure 3.3.1.

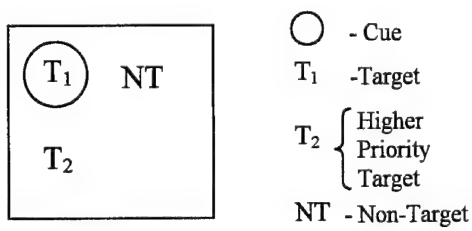


Figure 3.3.1. Illustration of a type A (attentional) costs. The cue directs attention to a target at the expense of missing or drawing attention away from a higher priority target.

The circle around T_1 (a target) represents the presentation of a valid cue that directs attention to that area and facilitates detection of that target. However, a second target, T_2, is also present and in this case, represents a higher priority target that is simultaneously present (but uncued) in the visual field. A type A cost results when the cue causes unwanted attentional tunneling on T_1, causing T_2 to be missed or undetected.

Yeh, Wickens and Seagull (1998) and Yeh & Wickens (1998) observed this type A cost in their research with HMDs. Their study evaluated the potential benefits for HMD use by Army personnel by measuring their performance on a target detection task. In a paradigm similar to that

which will be employed in the current thesis, the subjects searched for simulated tanks, landmines, nuclear devices and soldiers in an immersive 3-walled display environment, using a simulated HMD that presented fully reliable cueing on some targets and no cueing on other targets, including some targets of higher priority. Yeh et al. showed that cued targets were found significantly faster than uncued targets. However, when a target was cued, and a second target, which was uncued but considered a higher priority target, was present at the same time, the higher priority target was sometimes ignored in favor of reporting the lower priority target that was cued. Yeh et al. also found that the participant's field of view might even pass over the higher priority target and still not "see" the target because of the apparent tunneling effect of following the cue to the target. This narrowing of the attentional field could obviously have some serious consequences.

As discussed above, although cueing can aid target detection, there can be costs associated with cueing even when it is 100% reliable. However, as discussed below, when the cues are less than perfect, additional costs become apparent.

4. Cue Reliability

4.1 Costs of less than reliable cueing

As we have noted above, Fisher and Tan (1989) found that detection latencies were reduced when subjects received a valid cue that indicated where in the visual field the signal would occur. Similar to many of the previously mentioned studies, this orientation of attention to the location in space where a stimulus occurs significantly increased the efficiency of detection. However, Fisher and Tan (1989) also noted an inherent cost for cueing when the cue was *invalid*, that is, when the target stimulus occurred at a non-cued location. Cueing that was not correct resulted in slower reaction times than trials in which no cueing was presented at all. There was an inherent cost in the misallocation of attention.

The misallocation of attention caused by an invalid cue relates to the third property of attentional cueing: *reliability*. The reliability of an attentional cue is defined as the probability that the cue is valid or correct. Note that the reliability of a cue is quite different from the precision of a cue.

Fisher, Coury, Tengs, & Duffy (1989), also manipulated the level of highlighting validity by highlighting non-targets and by not highlighting targets. Their findings echoed those of Fisher and Tan (1989) showing the costs associated with cueing that is less than 100% reliable. As highlighting reliability or validity decreases, the benefits of highlighting are lost (Fisher and Tan, 1989).

The studies we reviewed above by Fisher and Tan (1989) and Fisher, Coury, Tengs, & Duffy (1989) both used small, rather well formatted, artificial displays of stimuli. Donner, O'Brien & Rudisill (1991) conducted a study that examined the benefits of highlighting validity in more complex alphanumeric displays. Subjects searched two Space Shuttle information displays

to answer questions about the value of a particular display item. In one experimental condition, subjects viewed the displays in their current format, which was described as poorly formatted, while in the other experimental condition the subjects viewed the displays in a reformatted version. Donner et al. (1989) reported similar findings to Fisher and Tan (1989) and Fisher, Coury, Tengs, & Duffy, (1989). When valid highlighting was present, subjects had faster response times on both the simple, reformatted displays and the more complex, poorly formatted displays than for the condition with no highlighting. However, when the highlighting validity was lowered to 50%, the benefits for highlighting were lost in both the reformatted and the poorly formatted displays, as subjects mean response times were no faster than in the non-highlighted condition.

From the review of cueing literature previously discussed, there is a cost associated with cueing that is invalid (Donner, et al., 1991; Fisher and Tan, 1989; Fisher, Coury, Tengs, & Duffy, 1989). All of these studies show an increase in response times when the attention is directed away from the target with cueing that was less than 100% valid or only partially reliable. There was not only the decrement in performance compared to the trials with valid cueing, but also a decrement in performance relative to the trials that used no cueing at all. This cost of partially reliable cueing will be labeled a type T (trust) cost. The type T cost is most evident in the examination of the cost/benefit tradeoff between when a cue is correct and when a cue is incorrect, as represented in figure 4.1.1.

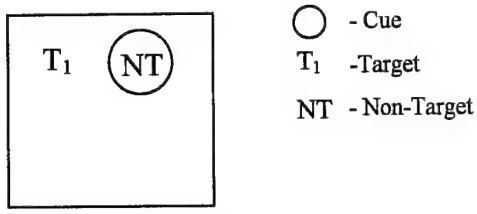


Figure 4.1.1. Illustration of a type T (trust) costs. The incorrect or non-valid cue directs attention to a non-target and it is classified as a target. If this also happens at the expense of missing the actual target(s), then the type T cost is coupled with a type A cost.

An example of a type T cost is illustrated in figure 4.1.1, when a non-target (NT) is cued (as represented by the circle around the term), attention is directed to that location and an action is taken, incorrectly, assuming that the non-target at the location is really a target. Potentially this trusted invalid cue also leads attentional resources away from other targets, a type A cost, resulting in targets either being missed or undetected.

An example of a Type T cost can be seen in an experiment conducted by Conejo and Wickens (1997). They manipulated the reliability of an automatic target cueing device in a simulated air-ground attack scenario. On some of the experimental trials, the cueing of the ground target by the automation was reliable whereas on other trials the automated cueing was not reliable (i.e., a non-target was cued). Although there was no indication of the reliability of the cueing for the subjects, there were environmental cues that could be used to manually confirm the cue's accuracy (or lack thereof). Conejo and Wickens found that when a non-target, or the wrong target, was cued, attention was directed to the highlighted cue and a decision was made to "attack" the non-target. In the short run, the costs of following cues when they are only partially reliable results in the tendency for a user to go "down the garden path" of automation trust, even when it should not be followed, the type T cost. In a combat environment, the soldier could fire at a cued 'target' when none is there (wasting ammunition), or the worse case of "the garden

path" phenomenon, a soldier could fire at a cued target when the cue defines enemy and it is really friendly (fratricide). In the long run, users might learn to distrust and ignore the automation cueing, "burned once, never to be burned again" (to the extent that this is possible since some cues (i.e. onsets) are harder to ignore). Once users begin to distrust or ignore the cueing, many of the benefits of the cueing will be completely lost.

From this discussion of type T costs that can be associated with automated cueing systems, the issues of trust, overtrust and mistrust of automation becomes a viable concern in the use of attention cueing information. A brief review on some of the literature of human trust in automation can present a better understanding of this trust/mistrust continuum.

4.2 Human Trust in Automation

Many studies have examined the degree of trust that humans display towards automation. This research provides excellent insight to the apparent trust, overtrust and mistrust of automation that users develop over time and experience with automation. Far from an all or none phenomenon, the degree of trust of automation by the human can be placed on a continuum (Parasuraman and Riley, 1997; Lee and Moray, 1991; Lee and Moray 1992; Lee and Moray, 1994; Muir and Moray, 1996; Singh, Molloy and Parasuraman, 1993). This trust continuum is directly related to the perceived reliability of the automation. The trust that a human has in a system is not always equal to the physical reliability of the automated system but to a concept of perceived automation reliability (Wickens, Gorden and Liu, 1998). User's expertise in the task that the automation is performing, user familiarity with, and user general affinity for automation all will directly or indirectly affect how the user trusts the automation (Riley, 1994; Lee and

Moray, 1992). Figure 4.2.1 shows the relationship between perceived and actual reliability and how a user whose trust is properly calibrated manifests neither overtrust or undertrust in a system.

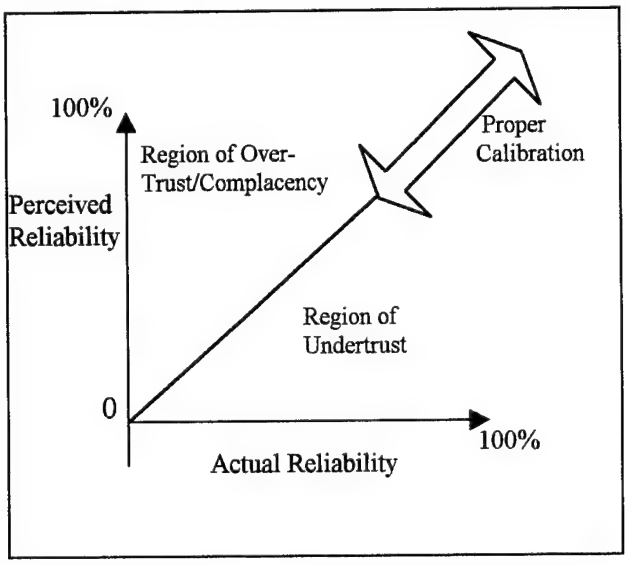


Figure 4.2.1. Reliability calibration.

Perceived and actual reliability can range from 0 to 100%. A user who has neither too much trust in the system nor too little, is said to be 'properly calibrated'. Figure 4.2.1 shows this region of proper calibration as the positive diagonal where the value of the X-axis equals the value of the Y-axis. A user whose trust can be placed along this line has the best opportunity to optimize the use of the automation. While the graph in figure 4.2.1 applies to most all forms and stages of automation, we will apply it in this thesis to the specific form of automation that pertains to attention cueing.

When the actual reliability of the system is low and the user places too much trust in the system, then the *Region of Overtrust/Complacency* is entered. When automation is accomplishing attention cueing, a user is very much susceptible to Type A and T costs discussed previously. On the other hand the user who perceives that the automation or system reliability is lower than the actual reliability operates in the *Region of Undertrust*. The problems associated

with this region are that the benefits that the automation can provide are not used most advantageously. A user who mistrusts the automation often times neglects or underutilizes the automation (Parasuraman and Riley, 1997). We will now review examples from the literature of both overtrust and undertrust in automation, as both will be examined in the current study.

Mosier et al (1998) explored pilot overtrust in automation exhibited by over-reliance of automated cues in the cockpit and labeled this an 'automation bias.' In their study, they had commercial pilots of automated aircraft fly a simulator with displays similar to the ones that could be found in a B747-400. In the simulated flight, an engine fire was reported by the automation. There were no other indicators in the engine parameters or any other indicators to suggest a fire and therefore confirm the diagnosis made by the automation. A significant number of the pilots not only trusted the automation and shut down the perfectly good engine, but also reported 'ghost memories' during a post flight interview; i.e., they believed that they had seen other indications of an engine fire, when indeed they could not have (no other indicators existed). The pilots had such a high level of expectancy of the abnormal conditions signaled by the automation that they perceived conditions that did not exist. This failure to question the automation and place an almost blind trust in it could obviously have serious consequences.

On the opposite end of the trust continuum, the perceived reliability of the automation could be low, causing the user's proper calibration to be skewed into the "Region of Undertrust, as shown in figure 4.2.1. This undertrust or mis-trust of the automation can lead to serious consequences as well. Sorkin's (1988) research with alarms presents excellent examples in user's undertrust of automation. He reports that alarms that tend to alert the user too many times with "false alarms", will lead users to ignore the alarms on occasions when indeed they are accurate (Sorkin, 1988).

Users of automation undergo changes in their trust of automation as a result of positive and negative experiences with it over time. Muir and Moray (1996) report that any sign of incompetence by the automation greatly reduces the trust the operator has in the system, even if the incompetence had no real visible detectable effects on the performance of the system. They also reported that although performance usually makes a rapid recovery after the user detects faults in the automation, a recovery of trust is not nearly as rapid. Lee and Moray (1992) reported that it took significant time and interaction with the automation performing reliably before the user would regain the levels of trust in the system that existed before the system failure.

In addition to the influence of actual failures on mistrust, Sarter and Woods (1995) show how users who do not have a full understanding of automation can come to mistrust the automation's actions. Their work with commercial airline pilots showed that pilots with a poor mental model of the planes' flight management system, along with inadequate cockpit displays, have mistrusted some of the aircraft's automation to a greater extent than is warranted.

Sometimes mistrust of automation may result simply because users have an inflated sense of their own competence. For example, Entin (1998) looked at the influence of different automation reliability levels of an aided target recognition (ATR) system on the user's performance and confidence in target detection decisions. Entin found that users of the ATR system had higher confidence in their own target decisions than that of the ATR system even when the ATR had high accuracy. Users of the ATR system failed to properly calibrate and found themselves operating in the region of undertrust, mistrusting the automation to a greater extent than was warranted. Similar results were found regarding an automated scheduling task examined by Liu, Fuld and Wickens (1993).

Users do not necessarily have to believe that the automation is 100% reliable for the automation to be useful. In a study by Kantowitz, Hanowski, and Kantowitz (1997), information reliability for a vehicle route guidance system was manipulated to see how the differing levels of reliability would affect subjects' acceptance of the automated system. The researchers varied the levels of accuracy of route guidance information at 100%, 71%, and 43%. The results showed that information that was 71% accurate was acceptable by the subjects and considered useful as well, whereas the information that was at 43% accuracy was not used. Research conducted by Gempler and Wickens (1998) looking at Cockpit Display of Traffic Information (CDTI) for pilots, showed that automation was useful even when the pilots knew it was only 87% reliable. This study will be reviewed in more depth in section 4.4.

We have presented many examples in the literature showing the difficulties in proper reliability calibration. Lee and Moray (1994) have suggested that there is a need to provide operators with better information regarding their performance, and the performance of the automation, so that the operator's self confidence and trust reflect true capabilities, and promote the appropriate use of automation. A better calibration of reliability might be gained through display design that explicitly displays the reliability of a system. If the user of some display/system is better able to implement a proper reliability calibration, he or she might know when it is necessary to allocate attention to different information sources or more information sources, beyond that which is cued or signaled by the automation.

In summary, there are several variables or characteristics that can lead to more or less allocation of attention to particular sources within a display. In a type A (attention) cost, certain characteristics of one channel will lead to more processing of that channel at the detriment of another channel (see figure 3.3.1). In a type T (trust) cost, certain characteristics of the channel

will lead it to be trusted more than is deserved on the basis on its actual reliability (see figure 4.1.1). What are these variables? Wickens, Pringle, & Merlo (1999) place these variables into two distinct categories: *implicit* variables and *explicit* variables, a distinction which we discuss in the following sections.

4.3 Implicit variables

Implicit variables are those properties that are inherent in a display channel and cause more or less processing or attention to be allocated to that channel by the user, even though they may not be explicitly designed or intended to do so by the display designer. An example of an implicit variable might be the location of the display item in a salient position (i.e., the top/front center) in the display panel. This feature would draw users' attention because of its prominent spatial location in reference to the user. Another example might be a large display item, which is more salient than the others around it, again drawing the users' attention. Ververs and Wickens (1998) compared many of these implicit variables like symbology saliency and spatial proximity and examined how these implicit variables affected divided and focused attention. Other facets of a display that might affect processing by the user are not yet as well validated by research. For example, a high level of realism of some display icon or presentation may increase the perceived reliability of the information to a degree that is significantly higher than is justified from the integrity of the data it represents (Theunissen, 1998). Highly realistic displays could be incorrectly processed as accurate data, consequently users may make strong inferences (seeing is believing) from less than accurate information. In these instances, we assume that such strong inferences represent the allocation of a disproportionate amount of attention to the display element in question.

As an example, Ockerman and Pritchett (1988) conducted an experiment examining the use of HMDs to assist pilots in preflight aircraft inspections. In their experiment, the control group of pilots used the traditional paper checklist to conduct the preflight inspection, whereas the two experimental groups used the aid of an HMD with either text instructions or real pictures showing the steps for the aircraft inspection. Their results indicated that the subjects using both the text only and graphic HMD tended to limit their inspection to only the things that were shown in the text or pictures and seemed to allow the "realism" of the display to limit their processing of the inspection steps. In both the text and picture conditions for the HMD, the subjects limited their inspection to only check the items that the display showed to check. The subjects using the paper checklist inspected these same items that the paper instructions said to check, but they were more thorough in their inspection than the HMD subjects. The subjects were essentially given the same information, just in a different display platform. Both the text and the pictures in the HMD seemed to cause the subjects attention to be guided exactly to what was written or pictured and nothing else. Subjects who had only a text version of the pre-inspection checklist and did not use the HMD did not show this apparent tunneling effect and found more faults in the inspection than did the subjects in the HMD condition.

While there were statistically significant differences in fault detection between the control subjects and the subjects wearing the HMD, there was no significant difference between the HMD picture and the HMD text groups. A factor that appears to implicitly modulate the amount of trust a user has in a display could be in the nature of the display platform. Yeh and Wickens (1998) compared the display of target cueing information between an HMD and a hand held display. As previously discussed, they found an apparent attentional tunneling that we labeled the

Type A (attention) cost. This Type A cost was reduced when the hand held display was used than when the HMD was used.

To summarize, certain features of the display, like the HMD platform, were not explicitly designed to cause users to process that information at the expense of other information; but this is what happened, and the effects of such features should not be overlooked.

4.4 Explicit variables

In contrast to implicit variables, explicit variables are usually part of a designer's choice to provide a way of informing the user that some sources are better, more reliable, or more important than others, and hence that these sources should be the focus of greater attention. Explicit variables can be separated into two different categories, those that focus/narrow attention, and those that broaden attention.. First there are explicit variables that deal with explicit attentional direction. Some examples of these types were discussed in section 3.2, like highlighting and attentional cueing, all examples of variables that are part of a designer's choice to direct attention and help focus the user's attention on relevant items.

The second category of explicit variables is those that explicitly display cueing precision or the level of source reliability and are designed to explicitly broaden attention. An example of an explicit variable of this type can be seen in the United States Army's standard map symbologies. The military has a standard set of symbols that it uses to portray enemy units/equipment on maps and terrain sketches. These symbologies are published in the Army Field Manual FM 101-5-1, Operational Terms and Symbols. If the precise location of an enemy unit is not certain or has not been confirmed, then the enemy's location is "templated," on the basis of map analysis, historical precedence, etc. The symbology of this enemy unit will be rendered just as it always is, but it will

be composed of dashed lines instead of solid ones. This dashed symbology means that the estimate of this symbology's spatial position or unit composition is not necessarily 100% reliable. The presence of this dashed symbology does not represent any specific calculated reliability of the information, it only serves to display that the information might be less than 100% reliable. The intent of this salient difference in the symbology presentation (solid vs. cued) is to explicitly show a difference in what is certain and uncertain about the enemy's position and or composition. To the extent that the user correctly infers the reduced precision of the dashed symbology, he will be explicitly encouraged to broaden the scope of attention to other locations where an enemy unit might be located.

Researchers have shown that explicitly displaying the reliability or differences in reliability can be beneficial. In the context of alarm uncertainty, Sorkin, Kantowitz, and Kantowitz (1988) showed that a likelihood alarm display (LAD) could be used to provide information about event likelihood. LADs are an elaboration of the traditional binary alarm, i.e. on or off. LADs communicate varying levels of alarm certainty that the critical condition actually exists. The 3-level alarm might display 'no danger' at one extreme, 'moderate probability of danger' at the middle of the range, and 'certain danger' at the far extreme. Sorkin, et. al. (1988) showed that LADs could improve attention allocation among several tasks and also provide information that is more easily integrated into a user's decisions.

Other researchers have also demonstrated the importance of explicitly displaying the reliability of information and have examined different ways of presenting it. Kirschenbaum and Arruda (1994) examined whether graphical and verbal representations of location uncertainty would have varying effects on decision making performance. In their study, experienced naval officers conducted a targeting task using information from a sonar display. The reliability of the

information was varied as a result of the effects of oceanic noise, etc. To show uncertainty, a 95% confidence ellipse was displayed for the spatial display of reliability, or an adjective describing verbal uncertainty was used in verbal trials. The results showed that officers using the spatial display of reliability had more accurate judgements.

Montgomery and Sorkin (1996) examined whether or not visual display conditions could aid observers in prioritizing information from display elements that differed in reliability. They used a visual display of reliability to help subjects detect whether the information that was conveyed by an array of nine graphical elements represented a signal or noise output. They found that subjects' detection performance and weighting efficiency were highest when the higher reliability of the source was cued by the greater luminance of the display element, as opposed to a condition in which source reliability information was not displayed.

Andre and Cutler (1998) examined the effects of displaying positional uncertainty on pilot's hazard awareness. Using a simulated navigational display on a PC with a 17-inch color monitor, they had subjects follow a flight path that had a "meteor" on the flight path that had to be avoided. Subjects were to stay as close to the flight path as possible but avoid colliding with the meteor. The position uncertainty of the meteor was varied over three levels described as: accurate, moderately uncertain, and highly uncertain. The researchers compared three uncertainty symbologies against a baseline condition. These three different symbologies included a *text condition*, which had a numeric value representing uncertainty located on the meteor, *a graphical-implicit condition*, in which color signaled the positional uncertainty of the meteor, and a *graphical-explicit condition*, in which variable radius circles signaled the positional uncertainty of the meteor. The baseline condition presented the meteor proceeding along the flight path with no indication of the reliability of its path.

When compared to the baseline condition, all three uncertainly symbologies reduced the number of collisions under conditions of position uncertainty, but the best overall performance was found in the graphical-explicit display. The authors reported that this graphical-explicit display of system uncertainty seemed to elicit proper calibration by the subjects and promoted the appropriate level of conservative behavior even under the highest level of position uncertainty.

Gempler and Wickens (1998) examined a way to explicitly display reliability in a Cockpit Display of Traffic Information (CDTI). The CDTI is a display that assists pilots in making decisions to avoid possible collisions with other aircraft. Unlike other CDTI assessments, they included occasional trials in which the display predicted flight path of an intruding aircraft was inaccurate. This unreliability appeared to be responsible for a type T cost. Pilots maneuvered according to the conflict predicted by the erroneous symbology and appeared to ignore the actual displayed location and trend of the traffic, which was reliable. In their experiment, Gempler and Wickens tested pilots in a simulator using a CDTI that had a single line predictor of the intruder aircraft's predicted future flight path and compared this with a CDTI that had a wedge shaped predictor that showed a 95% confidence interval of the intruder aircraft's future position. Their findings showed that the display of the reliability had no significant impact on reducing the costs of the invalid predictor.

In conclusion, from our review of the literature we have illustrated several different methods for displaying reliability. The research suggests that certain methods of displaying reliability or uncertainty allow the user to be better calibrated, and generally provide an increase in performance compared to conditions with no display of reliability or uncertainty (Andre & Cutler (1998); Montogomery and Sorkin (1996); Kirschenbaum and Arruda (1994). However, Gempler

and Wickens (1998) indicated that some techniques of displaying uncertainty are not always successful.

In examining reliability, it is important to remain cognizant of the fact that, independently of how it is displayed, the actual reliability of systems can be defined in many ways. For the purpose of our study however, we distinguish between two characteristics of reliability: *precision reliability* and *absolute correctness.*

4.5 Defining Reliability/Uncertainty

The precision reliability of information has to do with limitations on the spatial accuracy with which information can be collected or presented.. A sensor might be reliable in discriminating whether a target is present or not, but it might not be as accurate in depicting the precise location of a target. This reliability could be the result of a limitation in the sensor, a function of the sensor's calibration, or a restriction on the number of sensors activating on a particular target leading to reduced location certainty, etc. For example, the metal detector that is used in most airports simply cues the presence of metal on a person that walks through the sensor, but it does not tell whether the metal is located in the person's breast pocket or in his shoe. The information presented by the metal detector regarding the physical presence of metal is indeed reliable and has lowered the time of search tremendously since the sensor has at least greatly narrowed the search (i.e. to a single person). With low precision reliability, the information is reliable, but has limitations to its spatial accuracy. In section 3.2 we discussed this characteristic of reliability in the context of imprecise highlighting, as examined by Fisher and his colleagues.

The second category of reliability relates to the absolute correctness of the cue. If sensors receive false or incorrect data and/or the automation that is delivering the information to the

display fails, the actual cue presented by the automation could be incorrect, either cueing an object it believes to be the target when there is none, or failing to cue an existing target. The presentation of information in this category does not simply lack precision, it is all together wrong. In the context of the airport walk through metal detector, if the sensitivity of the metal detector is set too high, the detector could alert to sources of metal which are actually as benign as hinges on an artificial limb or metal snaps on trousers. When the metal detector cues at this level of sensitivity, the number of false alarms becomes too great to facilitate normal passenger throughput for an airport. Of course if sensitivity is set too low, another possibly more serious type of error results, a weapon could be successfully smuggled through the inspection. This type of reliability will be called *absolute correctness.* The costs caused by breakdowns in this category of reliability are those costs that are associated with type T costs and the consequences are automation over-trust and under-trust as described in section 4.2.

The way we defined absolute correctness makes this category of reliability more associated with catastrophic failures or the high cost of many false alarms. Precision reliability on the other hand is more frequently encountered. Many systems that operators use on a daily basis lack precision reliability. For example, the ground-to-air radar that are used by most airports lack a great deal of precision and cannot exactly pinpoint the location of an aircraft. One of the ways air traffic controllers deal with this imprecision is to give each aircraft a large area of protected space.

The registration accuracy or the required precision that a cue must have to its target to be effective depends largely on the accuracy that is expected by the observer. For example, a cue that only indicates whether a target is in front or behind of an observer could be as much as 89° off, and still be helpful if it eliminates half of the search field for the observer. Whereas, a cue that

is expected to be within 10° of the target would presumably not be very helpful if it was off by as much as 89°. Although no studies were found that directly manipulated cue precision as an independent variable to explore registration accuracy, some studies have indirectly examined the effects of cue precision. Posner (1978) looked at the effectiveness of an arrow presented at the point of fixation directing attention to targets on a CRT that were 0.5, 6.9, or 25 degrees away from the point of fixation. He found no difference in terms of response time for target detection among the three different distances. In contrast, Rizzolatti, et al (1987) reported that cueing in the center of the visual field was only effective if it was within 4 degrees of the target. It may be inferred that the studies differed in the conspicuity of the target and its visibility in the periphery. With the more conspicuous target employed by Posner, subjects may have been able to expand their focus of the attention spotlight. It is important to note that, both of these experiments as well as the bulk of related experiments in the basic attention literature used simple visual search paradigms and used short stimulus onset asynchronies (SOAs) of 100-300ms before the onset of the target, which may limit their generalizability into a more applied setting. Additionally, the experiments offer no insight as to whether an explicit display of the level of imprecision or registration accuracy would allow for a lower registration accuracy as the user would expect the decrement in precision.

5. Summary

In conclusion, a variety of studies have demonstrated the benefits of attention guidance, through either direct cueing to a particular target (Posner, 1978; Posner, Snyder and Davidson, 1980; Rizzolatti, Riggio, Dascola, and Umiltá, 1987; Yeh, Wickens and Seagull, 1998; Yeh and Wickens, 1998), through more indirect techniques such as highlighting, that restricts the search space to a reduced number of potential targets or a reduced area in which the target may appear (Brown, 1991; Donner, O'Brien & Rudisill, 1991; Drury and Clement, 1978; Fisher and Tan, 1989; Fisher, Coury, Tengs, & Duffy, 1989; Posner, Snyder and Davidson, 1980; Smith and Goodwin, 1972; Van Orden and DiVita, 1993; Hofer, Palen, and Possolo, 1993) or through implicit techniques, such as size or spatial location. Some of these studies have also examined the two types of costs, first, when either non-cued material is relevant but not attended (Yeh, et. al. 1998; Yeh and Wickens, 1998) or second, when cueing is simply incorrect, directing attention to a location where there is no target (Conejo and Wickens, 1997). The taxonomy proposed by Wickens, Pringle and Merlo (1999) describes these two costs as type A and type T, respectively. Furthermore, a still more restricted number of studies have examined how display properties can modulate the breadth of attention, or level of trust in automation guidance accordingly, to dilute these undesirable costs (e.g. Gempler and Wickens, 1998; Kirschenbaum and Arudda, 1994; Montgomery and Sorkin, 1996; Sorkin, Kantowitz and Kantowitz, 1998; Andre and Cutler, 1998), but no studies have characterized the costs in terms of A and T costs, a specific goal of the current study.

Of these studies, only one, Yeh and Wickens (1998), has examined these effects within the context of the search of naturalistic scenes, characteristic of the ground soldier's task, of interest in the current study, and this study will provide the paradigm used in the current investigation.

However, the present study extends beyond the investigation by Yeh and Wickens (1998) in several respects. First, Yeh and Wickens only examined type A costs, whereas the current study will examine both type A and T costs. Second, while Yeh and Wickens found that type A costs were modulated by the type of platform, or level of immersion of the display cueing device (HMD vs. HHD), (i.e. implicit techniques), they did not include in their design any variable explicitly designed to attenuate the costs, by "diffusing" attention. This variable is incorporated in the current study. Third, Yeh and Wickens did not examine the time course of attention breadth that might be induced by failures of automation; indeed few other studies in the literature have done so (e.g. Moray and Lee, 1994), and none within the context of the type A/type T taxonomy of cueing costs. Fourth, although Yeh and Wickens reported some important differences in type A costs between the HMD and HHD conditions, these differences could be attributable to two other, potentially confounding aspects of their study: the restricted visibility of the HHD and the considerable weight of the HMD. These potentially confounding differences are eliminated in the current experiment. We have improved the visibility of the hand held device and accounted for the differences in weight between wearing an HMD and using a hand held device.

Thus, the current experiment replicates many aspects of the Yeh and Wickens (1998) paradigm. Soldiers searched a field for both cued and uncued targets. Infrequently these were presented concurrently with a higher priority uncued target, in order to infer type A costs. Searches were conducted alternating between either an HMD or an HHD. Cueing was varied between blocks in terms of the level of spatial precision with which the cue designates the target (and hence the breadth of the search field), and this precision information was explicitly displayed by the format of the cueing signal. Our interest is in whether this display would serve to broaden the distribution of attention, and whether this greater breadth would be reflected in a reduced type

A cost. In the final block, the cue failed catastrophically, in order to induce at type T cost, and we examined the consequence of the resulting loss of trust on the subsequent search trials. Specifically we were interested in whether 'trust destroyed' in the cueing might lead to a subsequent loss in the effectiveness of target cues, as well as a resulting reduction in type A costs, as inferred by improved detection of the unexpected, uncued target.

6. Method

6.1 Subjects

Twenty subjects participated in the experiment, eighteen males and two females. The subjects were either active duty Army personnel, soldiers in the Illinois National Guard or cadets in the Reserve Officer Training Corp at the University of Illinois. All subjects had either self-reported uncorrected 20/20 vision or had corrected 20/20 vision in which case the subject wore glasses underneath the apparatus.

6.2 The Task

In order to build upon the results of previous research, the task and display format used in this experiment were either identical or very similar to the paradigm used by Yeh, Wickens and Seagull (1998) and Yeh and Wickens (1998). Subjects performed a series of tasks with the primary task being that of target detection. Subjects scanned the display in search of one of four target types (tank, soldier, land mine or a nuclear device).

Once the target was detected, the subject was required to respond by pressing the appropriate key on an input device. Next, if the target was a tank or a soldier, the subject had to identify the target as friend or foe (based on whether the target was facing left or right respectively, see figure 6.2.1). Finally, the subject was required to report the azimuth angle at which the target was located. Three of the target types were presented 90% of the time (30% each) and the fourth target type (nuclear device) was only presented 10% of the time. The fourth target type (nuclear device) was considered unexpected. Unlike the other target types, the nuclear device was never presented alone. It was always presented concurrently with either a tank or a soldier. The nuclear device was presented at a higher contrast ratio with the

background than the other three targets, thus making the appearance of the nuclear device more salient. Subjects were informed that a nuclear device took precedence over all other target types and should be reported over any other target that was simultaneously present. Subjects were never told which target would be present on a given trial.

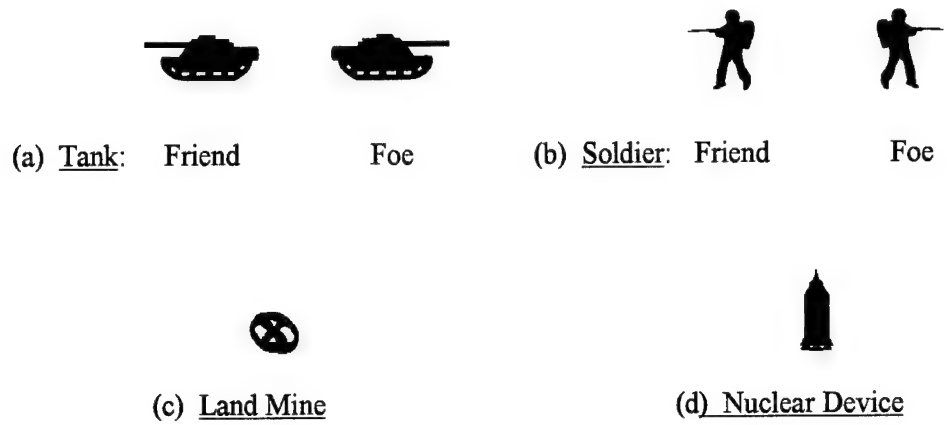


Figure 6.2.1 Targets. The targets that were used in the detection task.

The location of all tanks was cued with either an arrow pointing in the direction of the target based on the subject's current head position or a reticle in the case of the HHD. Only half of the land mines and soldiers were cued in such a manner.

While performing the target detection task, subjects also performed a secondary task. In the HMD this task was presented in the lower center of the display, so as not to obscure any of the targets, and in the HHD it was displayed on the center of the HHD, again not in the way of other pertinent information.

6.3 Apparatus and Materials

The terrain was the same as the terrain used in Yeh, Wickens and Seagull (1998), The terrain database was downloaded from the U.S. Geological Survey web site and it included data

from Austin, TX, Detroit, MI, and Jordan Valley, UT. The tanks, soldiers, and nuclear devices were camouflaged with shades of brown, green, and black whereas the land mines were presented in all black. The shading of the terrain varied so the intensity of the targets was adjusted at each location so that the contrast ratios between the target and the terrain were similar for all targets.

The terrain and targets were displayed on the walls of the Cave Automatic Virtual Environment (CAVE), a 10x10x9 foot room sized video environment. Subjects sat in the center of the room surrounded by the three walls. Targets were placed so that they appeared on one of the three walls: left, right or center. Targets were not placed close to corners or the floor since these areas could have some distortion due to the stark 90-degree angle created by the projection walls meeting each other or meeting the floor.

During the HMD portion of the experiment subjects wore head-tracked Crystaleyes shutter glasses, which only allowed the display symbology to be seen by the subjects' right eye, thus simulating a monocular transparent HMD. The head tracking insured that the display imagery was always in front of the subjects head. The HMD imagery was superimposed on the CAVE walls, moving correspodning to head rotation and was constrained to a field of view of 60 degrees laterally and vertically. The subjects' field of view of the far domain was not constrained by the shutter glasses. The subject could always see the far domain in his or her periphery completely unconstrained with both eyes.

During the HHD portion of the experiment subjects wore non-polarized sunglasses that matched the 30 % transmittance of the shutter glasses. This feature degraded the visibility of the far domain to a level equivalent to that viewed through the HMD, and hence equated visibility of the targets between the two displays. The HHD was a small portable 2.5" TV screen that

presented essentially the same information as the HMD. While conducting the searches in either the HMD or the HHD condition, the subjects wore an Army issue Kevlar helmet with a head tracker attached to the top center of the helmet to collect head movement in the X (left and right movement parallel to the horizon), Y (up and down movement perpendicular to the horizon), and Z (head cants) directions.

An example of both displays is presented in Figure 6.3.1, which depicts a sample of the terrain being searched, the heading information, the cueing information, and the secondary task.

(A)

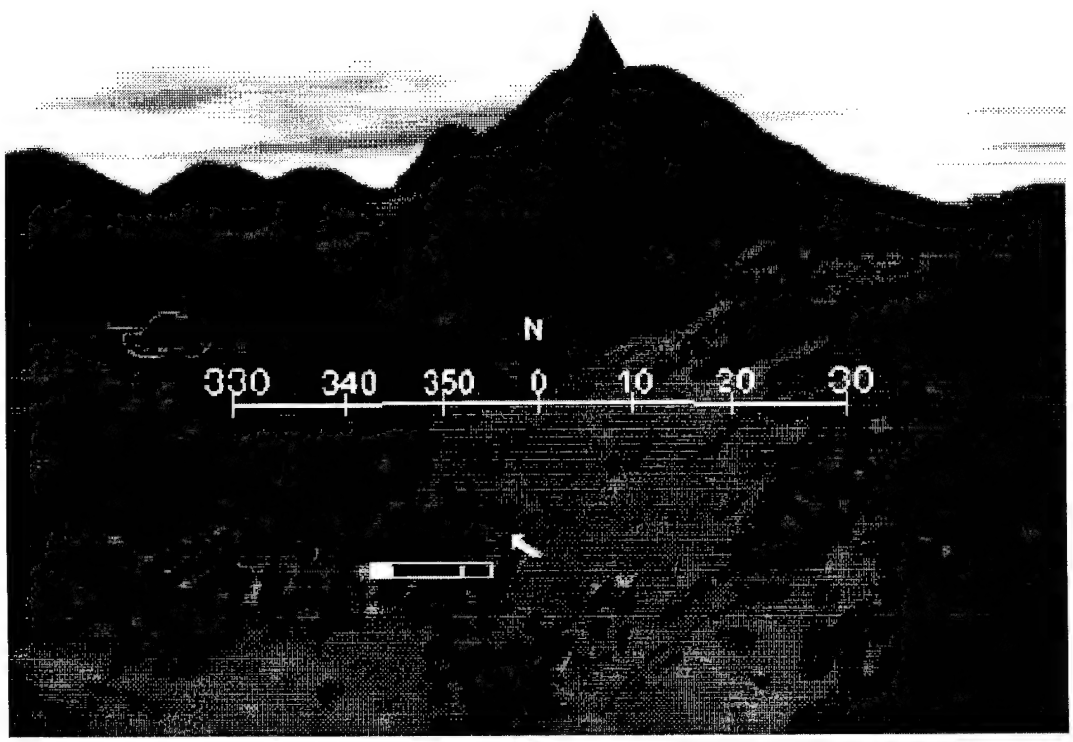


Figure 6.3.1. Displays – (a) HMD example. The HMD condition presents the heading, the cueing and the secondary task information superimposed onto the far domain. A target tank is shown at azimuth 325°. The cueing arrow that is shown will turn into a reticle as the subject rotates his or her head to bring the target into the FOV.

(B)

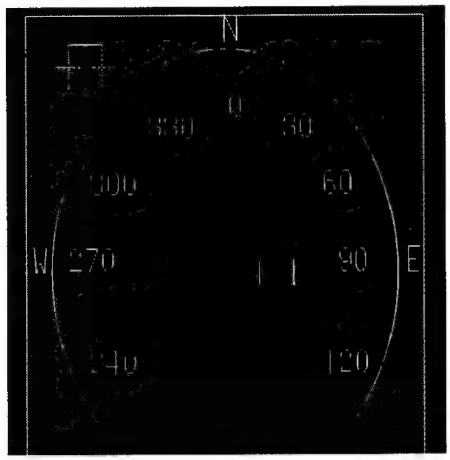


Figure 6.3.1. (continued) (b) HHD example. A cue shown at azimuth 325°. During the HHD condition the far domain is seen, but all of the additional information must be accessed from the HHD, only the target is located in the far domain.

On cued trials, a cue was presented to signal the current lateral and vertical location of a target with respect to the subject's current head orientation. Referring to part A of figure 6.3.1, to describe the effects of HMD cueing, the subject is looking into the center of the display. The cueing arrow is pointing left and up, which is in the general direction that the target is located. When the subject's head is moved to a position that places the target within 40 degrees of the center of his or her field of view, the arrow turns into a reticle and superimposes itself on top of the target. The reticle that forms over the target is identical to the reticle that is used for cueing target location in the HHD condition. This reticle can be seen in part B of figure 6.3.1. During trials that were displayed as containing precision errors, (see below) the cue, both in the arrow and reticle form for the HMD and the reticle form alone in the HHD, was composed of dashed lines instead of solid lines.

For the secondary task, subjects were told that their radio was having problems and that they would have to monitor a horizontal bar located in the bottom center of both types of

displays (see fig 6.3.1). This solid bar grew from left to right gradually and unpredictably, filling in a rectangle that contained it. There was a vertical marker located about 3/5 of the way down the rectangle denoting the point after which the subject could push a button to reset the growing solid bar. Subjects were informed that they must stay aware of the growing solid bar and never let it grow to the point where it touches the right side of the rectangle that contained it. If they tried to reset the rectangle before it reached the vertical marker, nothing would happen. Once subjects responded with the detection of the target, for the primary target detection task, the secondary task stopped. .

6.4 Experimental Design

The experiment used a within-subjects factorial design. The secondary task was present on all trials. The order of presentation of the different levels of cueing reliability and the viewing platform was counterbalanced between subjects.

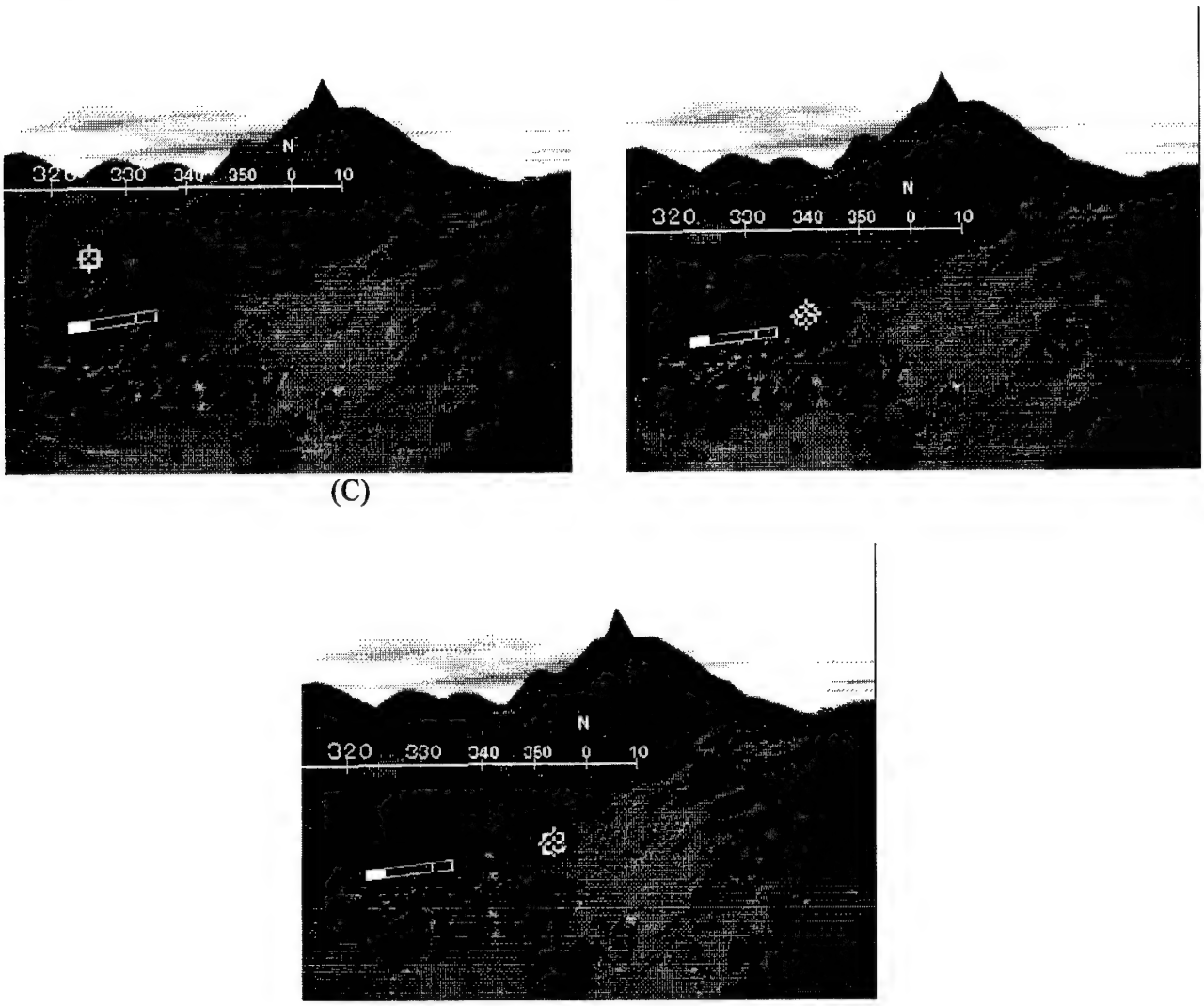
There were 11 blocks of trials. A block of trials consisted of 20 different target search scenarios. During a block each of the three main target types was presented alone six times (3 X 6 = 18 trials). Each target type was presented to each wall twice, once as a friendly, once as a foe. In addition to the 18 single target trials, an accurately cued soldier and tank were each presented on either the left or right wall with a nuclear device, to make a total of twenty scenes. The twenty scenes in a block used the same terrain scene throughout.

Subjects performed the first five blocks using either the HMD or the HHD, and then the second five blocks using the other display. After the subject completed the 10th block, they were told they would have one more block that would be similar to the last block completed. This 11th

block had trials that contained errors in precision well beyond that which were to be expected, representing a catastrophic failure of the system.

During each block, half (seven) of the cues were precise (solid lines) and the other half were imprecise (dashed lines) appearing in random sequence. The degree of imprecision alternated between blocks in a manner that will be described below. Of these 7 dashed cue trials, only 5 actually contained imprecision. The other 2 dashed cue trials contained no imprecision and placed the reticle right on the target. On one of the two precise cueing trials, a nuclear device was present concurrently with a tank or solider in order to measure the type A cost. On these trials the dashed cue was right on the soldier or the tank, leaving the higher priority target, the nuclear device, uncued. The other nuclear device for that block would be present with a tank or soldier that was cued with an arrow/reticle that was composed of solid lines (i.e. precise cueing). Thus the measurement of type A costs for both imprecise and precise cueing was done when the cue itself was on the target, so differences between symbology in the detectability of the device (assessing type A costs), could be attributed to the attention strategy induced by the symbology and not to the particular location of the expected target relative to the cue. Blocks alternated between those in which the dashed cues indicated small imprecision (45° diameter) and large imprecision (90° diameter). Figure 6.5.3 shows the three levels of proximity reliability with actual scenes from the experimental database. These pictures differ from the scene shown in figure 6.3.1 part a in that the follow pictures show the cueing arrow after it has turned into a reticle. Additionally, note that the reticles in figures b and c are shown in the dashed format.

(A) (B)



(C)

Figure 6.5.1 Differences in precision reliability. All three scenes have the same target located at 325 degrees and just below the azimuth scale. Depending on the trial block, the precision reliability can vary between 0 degrees (the cue is right on) to the cue being off as much as the extreme range given.– (a) Solid cue forming a reticle over the target ($0^\circ - 7.5^\circ$). (b) Dashed cue with degraded precision reliability ($7.5^\circ - 22.5^\circ$) off from the target. (c) Dashed cue with poor precision reliability ($22.5^\circ - 45^\circ$) off from the target.

6.5 Procedure

Subjects took approximately two hours to complete the experiment. Subjects were given a five-minute break between the change over of display platforms. Subjects were instructed to pretend that they were scouts and that they were to search for targets in unfamiliar territory.

They were to look for targets, identify them as friend or foe if relevant and report the azimuth to the target. Additionally, they were requested to monitor the aforementioned secondary task that represented a time sensitive radio problem that required a button push in conjunction with certain display parameters. This secondary task required relatively constant monitoring until the target was detected, at which point it was deactivated on the display, during the target identification and azimuth report phase.

Subjects interacted with the display using a wand and shutter glasses or a wand and a hand held device. The wand is shown in figure 6.5.1.

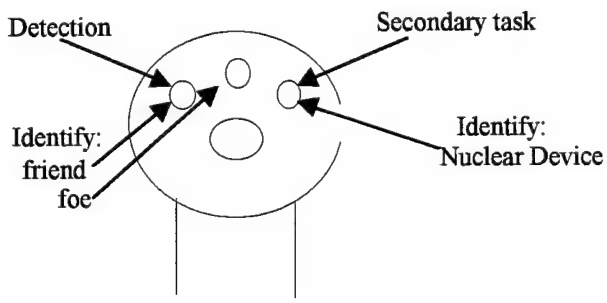


Figure 6.5.2. The upper portion of the wand. The three circles on the top of the wand are buttons activated by the thumb. The larger circle below the three buttons was a thumb-activated joystick not used in the present experiment.

Subjects responded to the different tasks using the wand's three spring-loaded buttons. They responded to the secondary task by pressing the button to the far right on the wand. Once the target was detected, subjects pressed the button to the far left on the wand. Once the detection of the target was made, the secondary task was stopped. After target detection occurred, subjects used the left and center button to classify the target as friend or foe respectively. If the target detected was a nuclear weapon, then, following the press of the detection button, they were instructed to press the far right button. Thus, although the far right button on the wand is the same one that was used for the secondary tasks, subjects knew that the secondary task had been

stopped after target detection and that the buttons on the wand had been remapped to the task of target identification (i.e. left-friend, center-foe, right-nuclear device). The button press for the target being friend or foe corresponding to the direction that the target was pointing, (eg. subjects pressed the left button if the tank or soldier was pointing left). Upon completion of the detection and identification the subjects verbally reported the relative location of the target by stating the target's azimuth bearing.

Once the experimenter had entered the reported heading the trial was complete and the display completely darkened. After the subject centered his or her direction of gaze (head) to a neutral starting position on the front wall, a new trial was initiated. Subjects completed 20 practice trials before the actual experimental trials were conducted. During the practice block subjects were able to ask any questions for clarification and all of the equipment was checked for proper fit and working condition. Before each block the subject was told the level of precision reliability of the different cues. In all of the experimental blocks subjects were informed that solid cues had excellent precision reliability with the cue forming a reticle around the target. Subjects were informed that dashed cues could be lacking in precision and a large foam circle was held up to the viewing screen to show subjects how far the dashed cue could be off in precision. This was done prior to each block in an effort to calibrate the subject's breadth of attention or search strategy to the level of imprecision.

As discussed previously, in each block the dashed cueing information was described to the subject as possessing one of 3 levels of imprecision: extremely reliable precision information, the cue being within ($0^\circ - 7.5^\circ$) from the target center; partially degraded precision reliability ($7.5^\circ - 22.5^\circ$ off from the target center); poor precision reliability ($22.5^\circ - 45^\circ$ off from the target center). Cues that were presented in a solid form were briefed to have good precision reliability,

and only dashed cues were used to signal partially degraded or poor precision reliability. Since levels of precision reliability were blocked, subjects knew what level of precision reliability to expect when a dashed cue was presented. Figure 6.5.2 illustrates the partially degraded and poor precision reliability conditions.

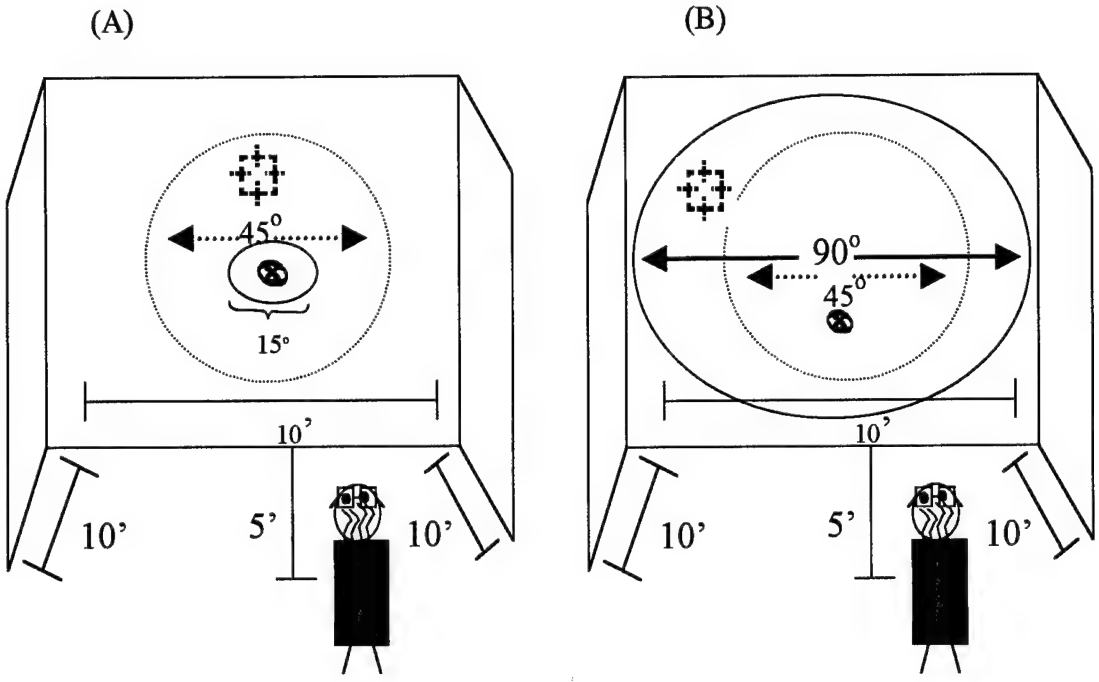


Figure 6.5.3 The different levels of cue imprecision. (A) Partially degraded reliability. A dashed cue (upper left) with partially degraded precision reliability (7.5° – 22.5°) off from the target (in this case a landmine located in the center of the viewing area.). The cueing arrow leads to a position in space within the 45° diameter circle. (B) Poor precision reliability. A dashed cue (far left) with poor precision reliability (22.5° – 45°) off from the target (landmine in the lower left). The cueing arrow leads to a position in space within the 90° diameter circle. This is approximately the visual angle subtended by a CAVE wall.

6.6 Performance Measures

The dependent variables that were collected from the primary target search task were response time and accuracy for target detection, target identification, and target heading. We also recorded the amount of head movement that occurred in the x-, y-, and z- axes. Also collected for each subject was the response time and accuracy from the secondary task. Since

the secondary task was dependent on an individual's time to detect the target (i.e. a target that is found in under 5 seconds would not require a response on the secondary task), the accuracy for the task was calculated as a proportion (rather than the actual number) of the number of hits to the number of total secondary tasks viewed.

7. Results

The data were examined to determine the effects of cueing, cue precision and of displaying cue precision level on target detection. Since there was a chance that subjects could mistake a terrain feature for an object, trials that had heading errors greater than $\pm 20^0$ were recorded as incorrect and were removed. Trials that had data that were greater than $\pm$ 3 standard deviations from the mean on any of the measured variables were also removed from the analysis. This occurred on no more that 3% of the trials. The data were analyzed for effects of counterbalancing order, either in order of display or in order of cue accuracy level to establish if either of these counterbalancing factors interacted with the display variables, in such a way to suggest that asymmetric transfer might have occurred. We found no statistically significant effects involving order. All results, therefore, will be reported collapsed across these order variables.

The data collected consisted of several dependent variables including the response time for target detection, identification and location of the target, as well as accuracy and speed of performance on the secondary task, and measures of head movement. Unless otherwise reported, all data were analyzed with a repeated measures within-subjects ANOVA using STATISTICA version 5.1. The analysis of the data is separated into the following sections:

7.1 The effects of spatially accurate, reliable cueing presentation
7.2 The effects of spatial precision on reliable cueing
7.3 Attention to expected vs. unexpected events (type A costs)
7.4 Head movement during target search
7.5 The effects of cueing on azimuth judgment
7.6 The effects of cueing on secondary task performance
7.7 Trust in automation during and after failures (type T costs): Block 11 data analysis

7.1 The effects of spatially accurate, reliable cueing

In order to determine the effect of cueing, a comparison of the detection times of cued versus uncued targets was conducted. Since tanks were always cued, they were not included in this analysis. Soldiers and landmines were the only two target types that were cued 50% of the time and uncued the other 50%. The data for soldiers and landmines were analyzed using a 2 (display platform: HMD or hand-held) x 2 (cue type: uncued or solidly cued) within subjects ANOVA. Figure 7.1.1 presents the effects of display and cueing on the two different target types. The bars in the figure show ± 1 standard error from the mean.

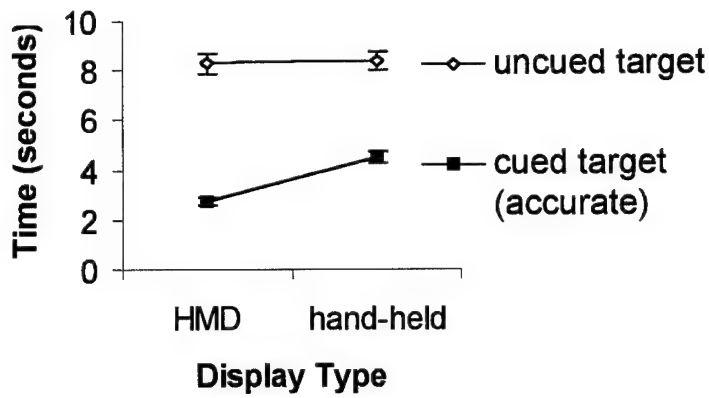


Figure 7.1.1 Detection times for both landmines and soldiers when target is cued or uncued.

As shown in figure 7.1.1, the benefits for reliable, spatially accurate cueing were consistent with our expectations. Cued soldiers and landmines were found faster than uncued soldiers and landmines, $F(1, 15) = 221.59$, $p < .01$. The data revealed that the two way interaction between display platform and cueing was significant, $F(1, 15) = 12.62$, $p = .003$. When the targets were uncued the display platform, HMD versus HHD, made no difference in the target detection time $F(1, 15) = .03$, $p = .868$. However, when the targets were cued, subjects wearing the HMD had an approximately 2 second faster detection latency, than when using the HHD, $F(1,15) = 83.63$, $p < .01$.

Because the salience of the two target types, landmines and soldiers, was quite different, we explored the possibilities of interactions between target type and the other independent variables. We conducted a 2 (display platform: HMD or hand-held) x 2 (cue type: uncued or solidly cued) x 2 (target type: soldier or landmine) within-subjects ANOVA. The detection times as a function of target type and display are presented in Figure 7.1.2, which presents the same data in figure 7.1.1, but now broken down by target type.

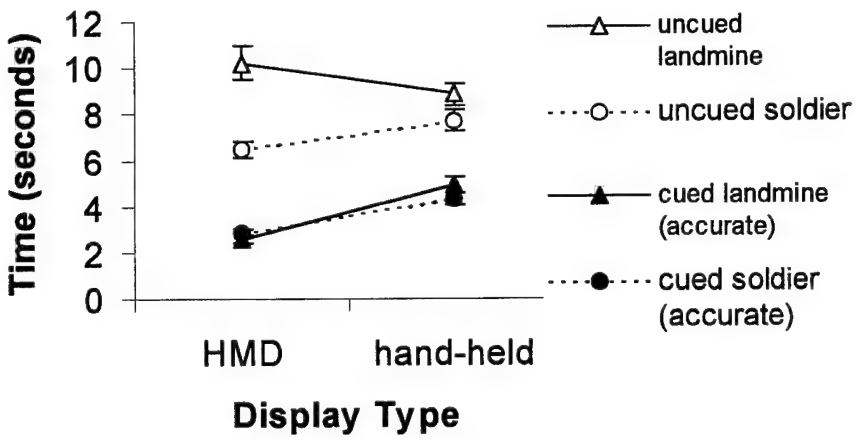


Figure 7.1.2 Detection times for landmines and soldiers when target is accurately cued or uncued.

As expected, regardless of the display platform, subjects showed the slowest detection times with the uncued landmines, the less salient target because of their much smaller visual angle (see figure 6.2.1 for a review of the different target types), $F(1, 15) = 22.32$, $p < .01$. Furthermore, independent of display platform, subjects' detection time for the landmine benefited more by the cueing than did their detection times of the more salient cued target, soldier, (cueing x target type interaction) $F(1, 15) = 16.53$, $p = .001$. There was also a significant three way interaction between display platform, target type and cueing, $F(1, 15) = 11.89$, $p = .004$. This interaction is explained by the fact that the subjects were faster detecting all cued targets (bottom two lines on

graph) while wearing an HMD than while using the HHD; however, when the targets were not cued (top two lines on graph) the display platform effect varied as a function of target type. The HMD helped the detection of the more salient soldier and hindered the detection of the less salient landmine. This two-way interaction, between display platform and target type for the uncued targets [$F(1, 15) = 7.42$, $p = .016$], can be explained in terms of a scan/clutter tradeoff. The clutter costs of the HMD dominates when the target is not salient (the landmines), thus the detection time of the uncued landmine is longest while wearing an HMD. On the other hand, the scan costs of the HHD dominates when the target is more salient, like the soldiers; thus the detection of the uncued soldiers was not as affected by clutter in the HMD and the longest detection time was for the HHD.

Since tanks were always cued, we did not include them in the above cueing benefit analysis. However, for reference we include the times below. Figure 7.1.3 shows the mean detect times for spatially accurate cued tanks, and analysis of these data revealed again that when targets were cued, the HMD provided the faster times of detection $F(1,15) = 11.23$, $p = .004$.

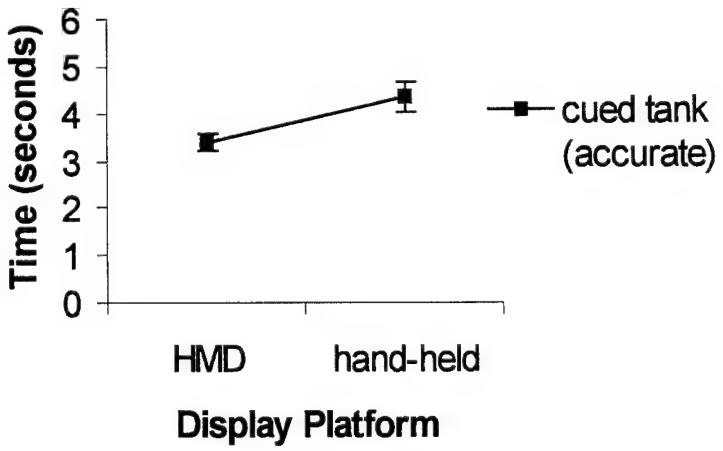


Figure 7.1.3 Detection times for accurately cued tanks.

Indeed, comparing the data in figure 7.1.3 with those in both of the bottom lines (cued targets) of 7.1.2, it appears that the use of HMDs for cued targets provides a consistent 1-2 second detection benefit over the HHD.

We manipulated whether the cue was presented dashed or solid (with a dashed presentation having the potential to be lacking some level of precision). Since some dashed cues were indeed spatially accurate in their target cueing precision (2 out of 7 per block), we wanted to determine if there was a difference in performance between the use of solid cues and the *accurate* dashed type cues. Therefore, we analyzed the data only containing accurately cued trials using a 2 (display) x 2 (cues: solid vs. dashed [with excellent precision accuracy]) x 3 (target: landmine, soldier, or tank) within subjects ANOVA. Since detection accuracy was perfect (soldiers, tanks, and landmines were always detected in all of the trials), the variance due to the experimental condition was examined exclusively by detection time. The results showed no effect for cue type across the two display types $F(1,15) = .227$, $p = .64$. This result was not anticipated in that we hypothesized that the dashed cue would cause a wider spotlight of attention or a wider zoom lens, thus a longer time to detect the target even when it was near the cue, (because of the postulated inverse relation between the size of the attentional field and processing times). With no apparent difference in latencies for cue type, the data suggest that subjects did not adopt a different attentional breadth for spatially accurate dashed cue trials, but rather adopted a scanning strategy that started at the cue location (thus benefiting detection of the two targets per block that appeared at that location) and worked outward from there. This search strategy will be discussed in section 7.4, as we report the analysis of head movement.

7.2 The effects of spatial precision in cueing

In order to determine the effects of spatial precision in cueing, a comparison of the detection times with the different cue types was conducted. This analysis compared detection times for targets with dashed cues when there was excellent spatial precision (i.e. the 2 out of 7 trials per block that dashed cues were spatially accurate), degraded precision reliability (45 degrees in diameter), and poor precision reliability (90 degrees in diameter), as well as targets that were uncued. The data were analyzed using a 2 (display platform) x 4 (cue type: accurate dashed cue, 45° off, 90° off or uncued) x 2 (target type: mine or soldier) within subjects ANOVA. Figure 7.2.4 shows the results for the target detection tasks.

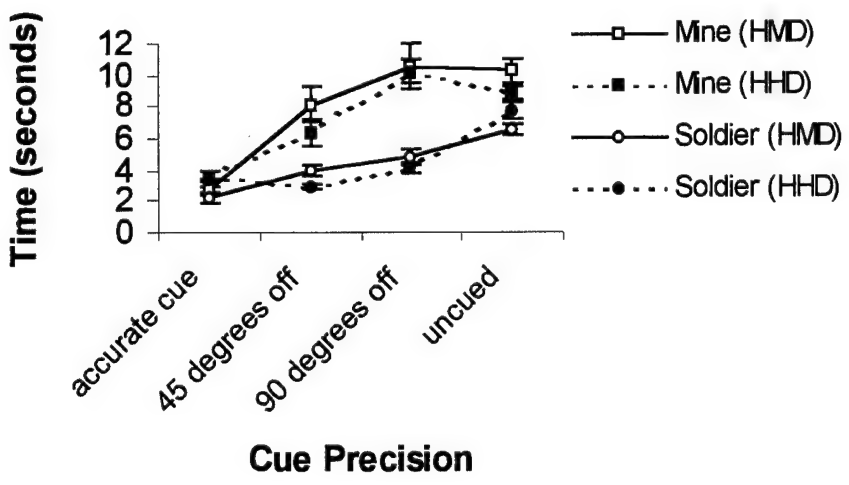


Figure 7.2.4. Effects of cue precision.

Not surprisingly, the spatially accurate dashed cues led to better detection performance than the dashed cues that contained degraded precision reliability. There was a main effect of cue precision $F(3, 45) = 47.42$, $p < .01$. This effect is also significant when the uncued trials are removed from the analysis, $F(2, 30) = 197.64$, $p < .01$. The performance decrement between the different cue precisions reflects significant differences at degraded precision condition (45° degrees off), $F(1,15) = 26.95$, $p < .01$, as well as at the poor precision reliability condition (90°

degrees off), $F(1,15) = 78.61$, $p = .000$. There was also a significant interaction between the target type and the cue precision $F(3,45) = 10.17$, $p = .000$, that can be seen in the loss of the cue precision hurting the detection of the targets with lower saliency (mines, top two lines on graph), to a greater extent than the targets with higher saliency (soldiers, bottom two lines). There was also a significant three-way interaction between the display platform, the target type and the cue precision $F(3,45) = 3.94$, $p = .01$. The soldier target type (high salience), in both platforms, benefited from cueing when the precision was degraded to the 90° condition. The landmine target type (low salience) did not benefit from the 90° cueing degradation condition, and landmines on the HHD actually suffered a cost from such imprecision, the detection latency increased (see figure 7.2.4) by almost 1 second.

7.3 Attention to expected vs. unexpected events (type A costs)

To explore the effects of cueing on the detection of expected (landmines, soldiers and tanks) versus unexpected targets (nuclear devices), a 2 (display platform: HMD or HHD) x 5 (cue type: solid and accurate, dashed and accurate, 45° precision, 90° precision, or uncued) x 3 (target type: soldier, tank, or nuclear device) within subjects ANOVA was conducted on the accuracy data from the target identification task, which are presented in figure 7.3.4. It will be recalled that the unexpected nuclear devices were always presented concurrently with an expected target, and that the expected target in this case was always cued accurately, with excellent spatial precision, even when the cue was presented dashed.

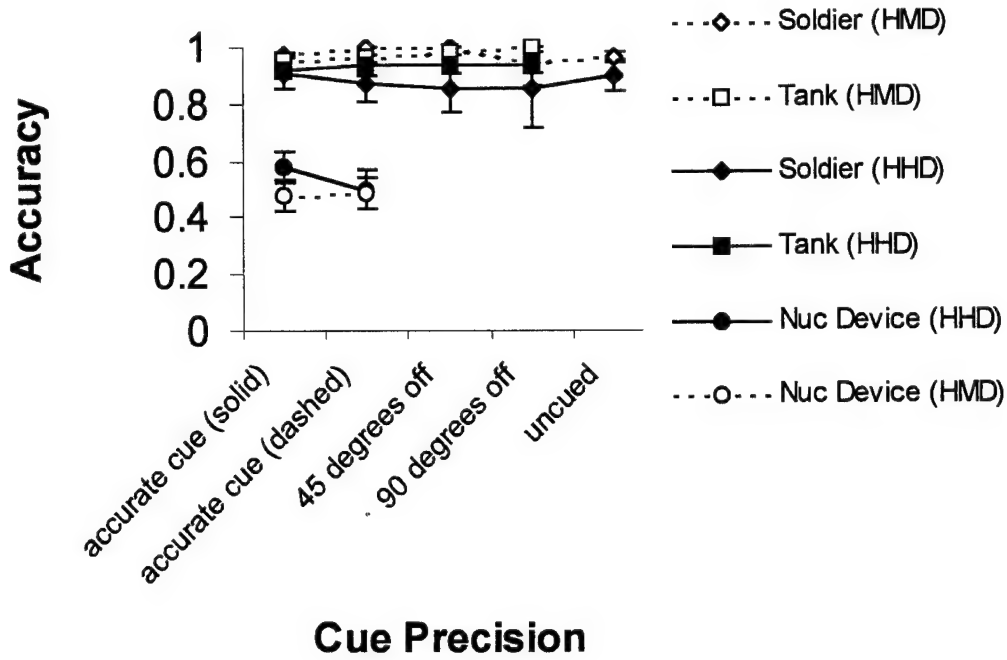


Figure 7.3.4 Subject accuracy for the target identification task.

Detection of the more expected targets was of high accuracy (averaging approximately 95%) and was not significantly affected by the different display types or cue types, as shown by the lines at the top of the figure. However, the unexpected but higher priority target, the nuclear device, had a significantly lower accuracy when compared to the other targets $F(2,30) = 110.20$, $p < .01$, as seen with the two shorter lines at the lower left in figure 7.3.4. These targets were never presented with imprecise cueing and were never cued directly. This finding of a low detection rate for the nuclear device was anticipated, as it replicates the previous studies (Yeh, Wickens, and Seagull, in press). (note: Yeh et al. compared the accuracy of detection of the unexpected event between trials where the simultaneously presented expected target was cued or uncued, whereas, we always presented cues for both occurrences of the nuclear device during the experimental block. The accuracy level we obtained: approximately 50%--corresponds quite closely to the level obtained for the corresponding condition used by Yeh et al.) It appears that

subjects followed the cue to the lower priority target, at the expense of missing the higher priority but uncued target (type A cost).

This finding of the type A cost was expected and we further predicted that two of our independent variables, the dashed cueing and the display platform, would mediate the Type A cost. However, consistent with our analysis of latencies in target detection, the different accurate cues (solid and dashed), resulted in no difference in target detection performance for the nuclear device $F(1,15) = .36$, $p = .56$. There was also no significant difference between the accuracy using the two different display platforms, $F(1,15) = .86$, $p = .34$. This finding of equivalent attentional tunneling by subjects regardless of viewing platform was not anticipated in light of previous findings that the HMD had produced greater tunneling (Yeh et al., in press). However, there was a non-significant trend toward an interaction between display platform and cue type $F(2,30) = 2.35$, $p = .15$, showing that the display platform did have some effect in mediating the type A cost. That is, as shown by the two lines at the bottom of figure 7.3.4, when cueing accuracy was certain (solid), the HMD supported less detection (greater type A costs) than did the HHD, a direction of effect consistent with that observed by Yeh et al.. When cueing accuracy was less certain (dashed cues), this cost disappeared.

7.4 Head movement during target search

In order to determine the effect of cueing on head movement during visual search, a comparison of the head movement prior to the detection of cued versus uncued targets was conducted. We first analyzed the data using a 2 (display) x 2 (accurate cue type: cued or cue dashed) within subjects ANOVA. Figure 7.4.4 shows the results for the average head movement during the target detection task along the x and y axes for the different cue types, dashed and

solid (the dashed cued trials in this analysis were restricted to the two trials per block when the target was precisely cued).

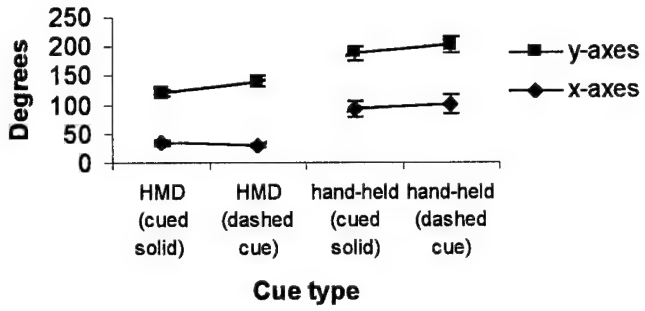


Figure 7.4.4 Effects of cueing on head movement.

Data regarding the head movement along the two axes show a significantly greater movement in both the x and y directions for the hand-held display [X-axes: $F(1,15) = 41.26$, $p < .01$; Y-axes: $F(1,15) = 31.98$, $p < .01$], hence helping to explain the cost of this display for cued target detection time, as discussed in section 7.1. However, consistent with the lack of difference in the target detection latencies as described in section 7.1, there was no difference in head movement between the two spatially accurate cue types (dashed and solid cues with excellent precision reliability): [X-axes: $F(1,15) = .62$, $p = .44$; Y-axes: $F(1,15) = 3.35$, $p = .087$].

To determine if the presence of cueing caused difference in the amount of head movement, a 2 (display platform) x 2 (cue type: cued or uncued) within subjects ANOVA was performed. As anticipated, there was significantly less head movement when the target was cued: [X-axes: $F(1,15) = 142.87$, $p < .01$; Y-axes: $F(1,15) = 59.25$, $p < .01$.] To determine if the precision of the cueing caused differences in the amount of head movement, a 2 (display platform) x 4 (cue precision: dashed cue containing accurate precision, 45° precision, 90° precision, or uncued) within subjects ANOVA was conducted. Figure 7.4.5 shows the results of this analysis.

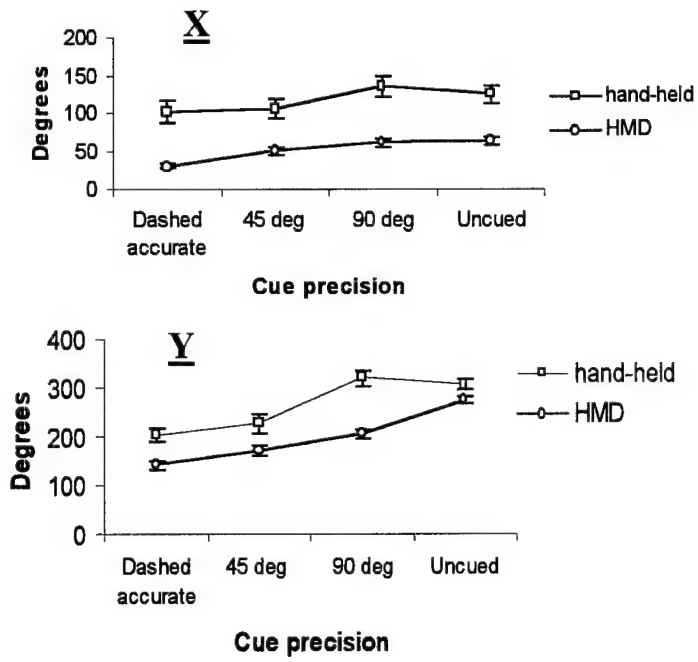


Figure 7.4.5. The effects of display and cue precision on head movement along the x and y axes.

Not surprisingly, there was a significant increase in the amount of head movement along both display types as the cue precision decreased: [X-axes: $F(1,15) = 28.24$, $p < .01$; Y-axes: $F(1,15) = 41.15$, $p < .01$. Note, as the cue precision reaches the 90° precision condition for the hand-held device, there is no difference in the amount of head movement between the cued and an uncued condition (the last two connected points on the above graph), [X-axes: $F(1,15) = 1.42$, $p = .25$; Y-axes: $F(1,15) = .76$, $p = .40$]. There was a marginally significant interaction between the display platform and the cue precision in the x-axis, and a significant interaction in the y -axis, [X-axes: $F(2, 30) = 3.07$, $p = .06$; Y-axes: $F(2,30) = 5.47$, $p = .01$]. As shown in the figure, this interaction suggests that the decrement in cue precision at the 90° condition caused a greater amount of head movement, particularly in the y-axis for the HHD. This interpretation is consistent with the other head movement results in that the hand-held display, even when accurate, does not give descriptive cueing information along the y-axis and therefore the subjects using the HHD had to conduct a more thorough search along that axis (vertical head movements) before finding the target

7.5 The effects of cueing on azimuth judgment

To determine the effects that display type and cueing have on azimuth judgments a 2 (display platform) x 2 (cue type: accurately cued vs. uncued) within subjects ANOVA was performed. Additionally, to determine if cue accuracy had any affects on azimuth judgments a 2 (display platform) x 3 (cue precision: dashed accurate, 45° precision or 90° precision) within subjects ANOVA was performed. The results of the analysis are shown below in figure 7.5.6.

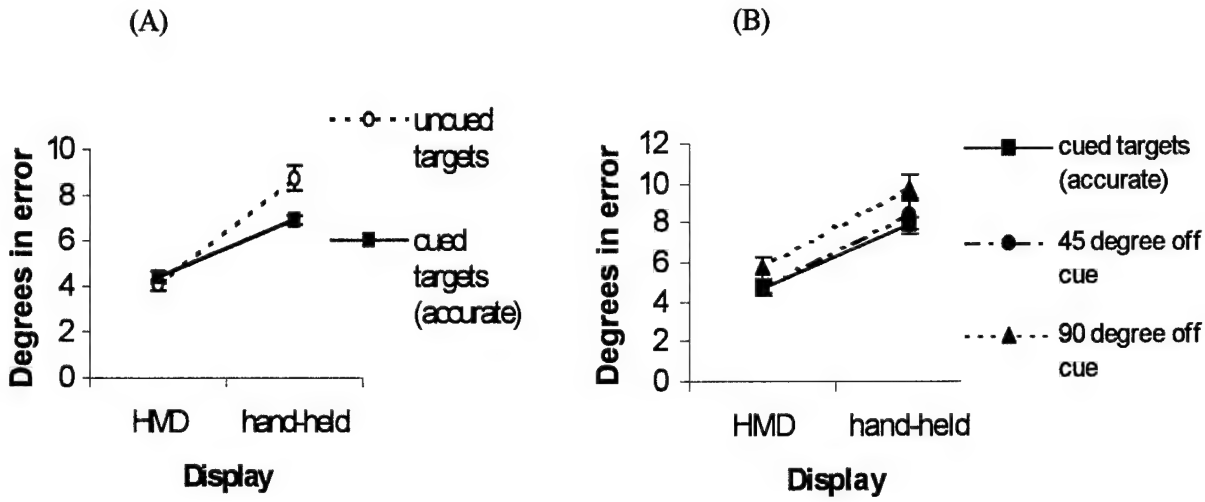


Figure 7.5.6. (a): The effects of display and cueing on azimuth judgements. (b): The effects of display and cue precision on azimuth judgment.

The analysis showed that users made more accurate responses (i.e. lower error) on target azimuth while wearing the HMD, $F(1,15) = 51.88$, $p < .01$. There was also a marginally significant benefit to azimuth judgement when the target was cued $F(1,15) = 8.75$, $p = .10$. However, there was a significant interaction between display type and cueing, $F(1,15) = 15.7$, $p = .001$ showing that this benefit for cueing was only present for the hand-held device. These effects can be interpreted in terms of cue resolution. First, there is greater spatial resolution of the azimuth scale presented on the HMD. The HMD contained world-referenced conformal imagery for the presentation of azimuth information (i.e. a 1-to-1 mapping in the far domain), whereas; the HHD azimuth information contained 240 degrees of measurements compressed into the small 2.5 inch

radius of the hand-held screen. Additionally, since the azimuth information is located in the near domain for the HHD (head down), when the target was uncued, eye and head movements were required between the target in the FFOV (far domain) and the azimuth scale on the display, thus accounting for the higher error in azimuth judgment for the HHD (as shown in figure 7.5.6 part a). There was also a main effect of precision, $F(1,15) = 4.5$, $p = .02$ (shown in figure 7.5.6, part b). When the cue was present and imprecise, we assume that subjects sometimes used the imprecise cue location rather than the actual target location, accounting for the significant greater error of the reported azimuths as the cue imprecision increased,

7.6 The effects of cueing on secondary task performance

To determine the effects of display platform and cueing on the secondary task a 2 (display) x 3 (cue: cued, uncued, or dashed cue) within subjects ANOVA was conducted on both the response time and the accuracy of the secondary task. The dashed cue trials collapse across all three dashed cue precision conditions (accurate, 45° precision, or 90° precision). The results of this analysis are shown in Figure 7.6.7.

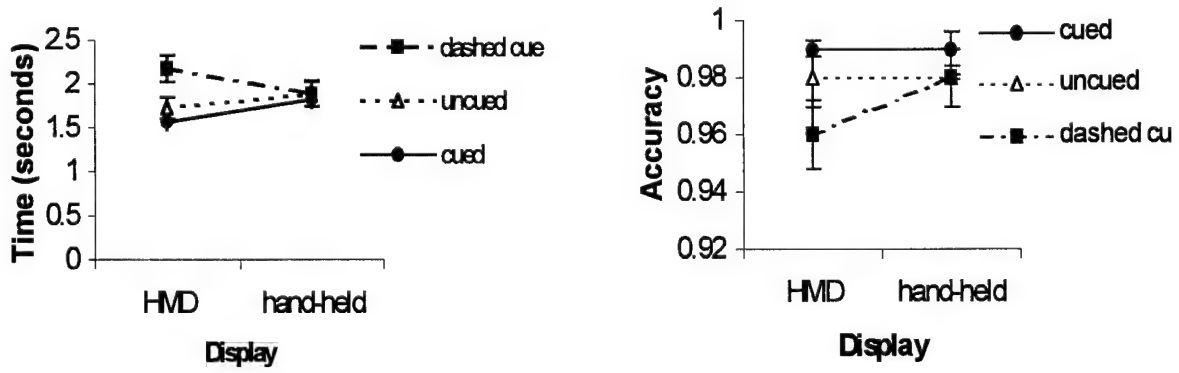


Figure 7.6.7. The effects of display type and cueing on secondary task performance.

The response time data showed no significant effect for display, $F(1,15) = .20$, $p = .66$. However, there was a significant increase in reaction time for the secondary task when the cues showed reduction of precision, $F(2,30) = 8.62$, $p = .001$. This increase is selective to the HMD condition, as shown by the significant interaction between displays and cue types, $F(2,30) = 6.81$, $p = .004$.

The accuracy data showed similar trends, with no significant effect for display, $F(1,15) = 2.88$, $p = .11$. This finding of equivalent performance in the secondary task between platforms was not anticipated. Yeh and Wickens (1998) found an advantage for the HMD in both time and accuracy. We think that the elimination of two potential confounds -- differences in subject head weight between the two display platforms and increasing the visibility of the HHD -- accounts for the differences in our results and theirs.

As with detection time, there was again a decrease in secondary task accuracy with the decrease in cue precision, as the effect for cue type was significant, $F(2,30) = 3.87$, $p = .03$. The interaction between the display and the cue type was marginally significant, $F(2,30) = 2.78$, $p = .078$, suggesting again that only the HMD performance was hampered by the decrease in cue precision.

7.7 Trust in automation during and after failures (type T costs): Block 11 data analysis

To explore the effects of trust, over-trust and mistrust in the cueing automation, we placed several catastrophic cueing errors in the last block of search trials. As previously discussed, these errors were not just lacking in precision, they were simply wrong, thereby lacking in the category of reliability we labeled "absolute correctness" (see section 4.4). This 11th block consisted of trials similar to those that the subjects had already completed; in fact the first three trials were normal uncued, cued solid and dashed cued trials, with the cued trials

containing no decrement in the cue precision. By the time the subject reached the 4th trial in block 11, he or she had experienced over 230 trials where the automation (the cueing), behaved exactly as the experimenter had briefed, allowing the subjects to develop some degree of trust in the target cueing. On the 4th trial of block 11, attention was directed to a position by a solid (high precision) cue that was 150 degrees from the actual target (a landmine). This cued location was an open space that was very unlikely to be confused for a target. All of the subjects presumably attended to the cued location, as their heads moved to the cued location.

Two subjects in the HMD condition immediately reported the detection of a target and made the appropriate response for a landmine, with one of the two subjects reporting the heading to the 'ghost landmine' and immediately centering his head for another trial as if no mistake had been made. The other subject verbally reported that he must have made an error, and that he could no longer see the landmine that he reported detecting and was not able give the azimuth to the target. The rest of the subjects noticed that the target was not present in the cued location, but continued to look in the immediate vicinity of the cued location. After searches averaging over 90 seconds in the HHD condition and 50 seconds in the HMD condition, subjects abandoned the cued location and went to other areas of the display to search for the target. The target was eventually found with an average detection time of 101.4 seconds for the HHD and 56.8 seconds for the HMD, a difference between the two display that was significant, $t(7) = -3.58$, $p = .006$. This dramatic effect of an unreliable and incorrect cue is shown in trial 4, in figure 7.7.8 (off the page!!), which shows the average detection times for each trial in block 11. The figure further illustrates the subsequent effects of this loss of trust on attention, which we now describe.

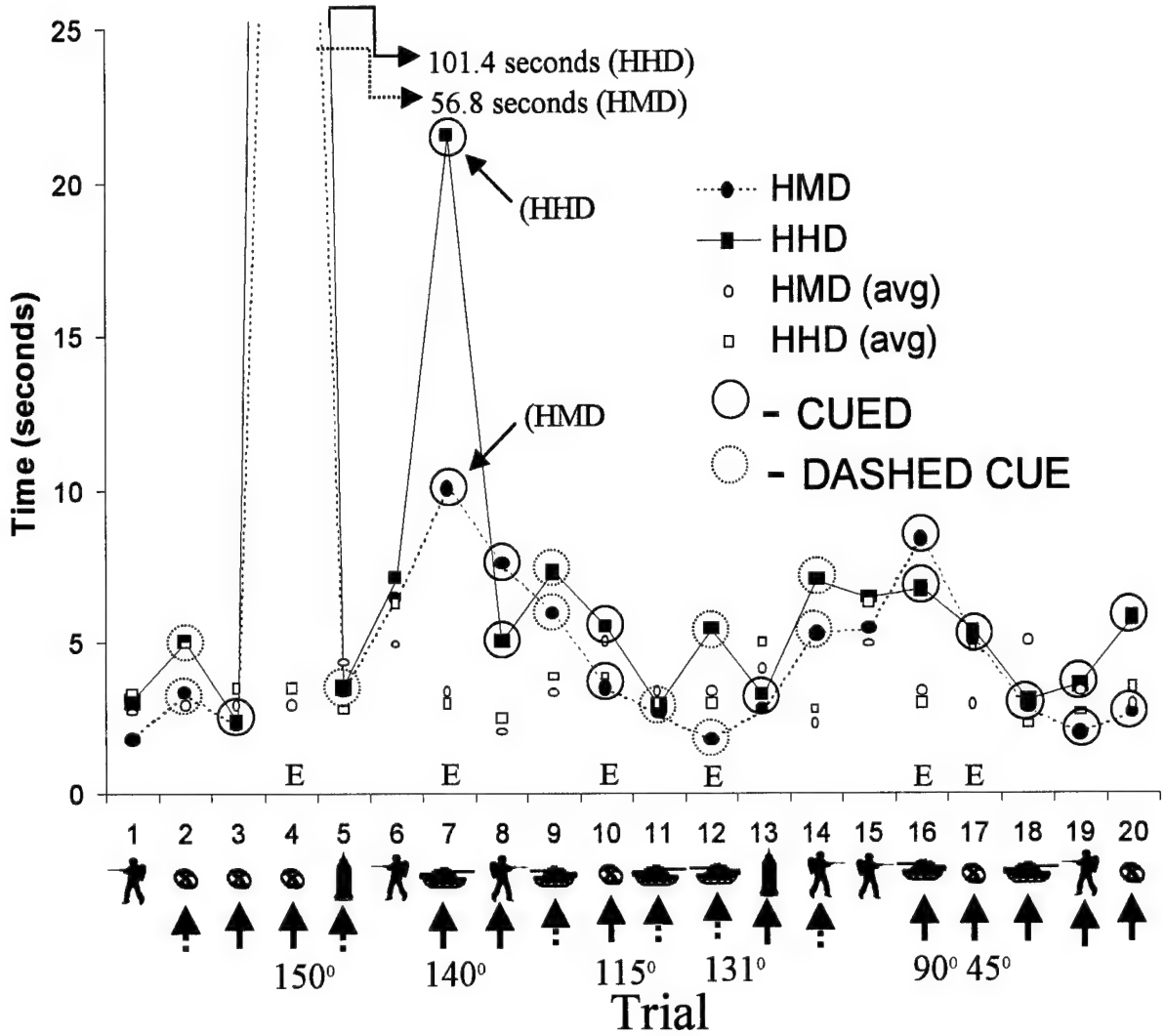


Figure 7.7.8. The results of trials in block 11 where the automation contains catastrophic failures. The target types for each trial are directly below the trial number. If cueing was present, the solid or dashed arrow directly beneath the target denotes the form of the cueing. A number beneath a cue is the number of degrees that the cued location is from the actual target center. The letter E represents the erroneous trials. The 2 lines (HMD and HHD) connect the detection time points observed in this unreliable block (block 11). The unconnected points on each trial represent the mean detection times on equivalent trials in previous blocks, before the cueing became unreliable.

Immediately following the 1st trial that contained erroneous cueing (trial 4, on which the long detection time can be seen in the figure), a trial with a cued target and an uncued nuclear device simultaneously present was presented to the subject (trial 5). While our previous analysis (see section 7.3) showed that the type A cost for cueing resulted in only about 50% nuclear device detection, subjects showed a much higher detection rate for the unexpected event right after the automation failure, with a 100% detection rate while using the HHD and an 82%

detection rate while using the HMD. On this trial, subjects quickly followed the cue to the cued target, but they did not seem to display the attentional tunneling that was so apparent in the previous like trials (measured by the detection accuracy of the unexpected target, nuclear device). A t-test was performed comparing the detection accuracy before the automation failure, and the accuracy immediately after the automation failure on corresponding nuclear device trials. There was a significant increase in the detection rate between the two HHD conditions (before and after the failure block), $t(7) = 5.23$, $p = .001$, and a marginally significant increase between the two HMD conditions (again, before and after the failure condition), $t(8) = 1.32$, $p = .22$ (note: we did not include the two subjects that reported the 'ghost' targets in this analysis and one subject did not complete block 11 because of a computer failure. Additionally, we used both of the nuclear device trials for each subject from block 11 to complete the t-test). Thus, the appearance of the failure seemed to "diffuse" attention and neutralize the type A cost.

We did not present another erroneous or invalid trial until trial number 7. On this trial we once again presented catastrophically erred cues with the precise (solid) cueing symbology so that it cued a location that was 140 degrees from the target location. Although subjects followed the solid cue to the erroneous location, no 'ghost targets' were reported nor did the subjects dwell in the general cued location as they had in the first failure. Subjects left the cued area and detected the target in half of the time that was required during the first failure, with the average detection time of the HHD being 21.5 seconds and the average time for the HMD being 10 seconds. These times are still somewhat larger than the 8 second search times for uncued trials as measured in previous blocks (see figure 7.1.1), and still show significant difference for display platform, $t(7) = -2.49$, $p = .04$.

By the time that the subjects encountered trial number 8, the automation had displayed erroneous cueing on two separate occasions. The cost for these automation failures started to reduce the trust in the cueing, as the 8th and 9th trials showed longer detection latencies (6 seconds) then the average cued latencies of the earlier blocks, as shown by the unconnected points beneath the lines in figure 7.7.8 (3.5 seconds), even though both of the targets in trials 8 and 9 were accurately cued (see figure 7.7.8). By trial 10 the presentation of erroneous cueing seemed to only slow the subject's detection time by the amount of time that is wasted in going to the erroneous cued location, as the user immediately disregarded the cue and found the target unaided. Thus we observed subject's following the cue to a position in space, and then if no target is present, initiating a global search that was similar to the searches that were conducted when no cueing was present i.e., they did not dwell longer around the cued region than any other region..

We presented another trial (trial 13) that had an accurately cued target with an uncued, but higher priority nuclear target present simultaneously. We again found detection accuracy that was higher than that recorded in the blocks prior to the first catastrophic failure in the automation, with the HHD having 100% detection and the HMD 82%. Thus the type A cost was reduced or eliminated following the demonstrated unreliability of the cue.

To examine the possible restoration of trust, we presented three more erroneous cues after trial 10. Here the benefits of cueing can be examined in either of two ways. First, one can compare cued detection times with their counterparts in earlier blocks before the trust had been betrayed (the unconnected points). Here trust appears to be fully restored by the final 3 trials shown in figure 7.7.8. Alternatively, one can compare the detection times with the single uncued soldier trials seen in trial numbers 1, 6, and 15. Here, although a modest cueing cost appears on

trial 16, by the last three trials, there is again a benefit. Thus subjects are apparently still benefiting from the cue, even as they have 'learned' that it is no longer 100% reliable.

8. Discussion

The current experiment was conducted to examine the effects of display platform, cueing, and the explicit display of cue precision reliability on type A and T costs. Additionally, we examined the effects of cue precision, as well as the time course of trust in automation (rendered here by automated attention cueing) after automation failures, and the effect of mistrust on the subsequent use of automation. While most of our findings are consistent with previous research, we found some results that were contrary to what we expected, and others that added important knowledge to the database linking visual attention deployment to automation guidance. We will now offer some explanations of our findings.

The use of the HMD, particularly with reliable and precise cueing, helped subjects in our target detection paradigm in a manner consistent with our review of the literature on head-up displays (HUDs) presented in section 2 (e.g., Wickens and Long, 1995; Osgood and Wells, 1991). By reducing head movement (scanning), the HMD allowed subjects to demonstrate superior performance (i.e. faster) in the target detection latencies of accurately cued targets. The HMD also helped with target azimuth accuracy because of the increased resolution offered by an ego-referenced azimuth scale. However, there was some evidence for HMD clutter costs with non-salient, uncued targets (i.e., the landmines, subtending a small visual angle; see figure 7.1.4.) When targets were uncued, there was a difference in the detection latencies for the different display platforms by target type, which was explained as a clutter/scan tradeoff (see section 7.1). Still, consistent with Fadden, Ververs, and Wickens (1998) findings in their meta-analysis comparing heads-up versus heads-down displays, the benefits of less scanning generally outweighed the costs of clutter in this experiment, when the HMD was compared with the hand-held display.

Not surprisingly, and replicating effects reported in Yeh, Wickens, and Seagull, (in press), accurately cued targets were found faster than uncued targets, (e.g., cued soldiers and landmines were found faster than uncued soldiers and landmines), with the fastest average detection times and greatest cueing benefits observed while subjects were wearing the HMD. The reasons that cueing provided better performance in the HMD platform are twofold. First, the ego-referenced HMD cueing was more precise in presenting the target location since the accurate cueing was usually superimposed directly on top of the target, thus providing the exact x and y location of the target, while the HHD lacked information concerning the target location in the y-axis. Secondly, since the cueing symbology was presented heads-up in the HMD, the head movement between the target and the cue symbology needed in the case of the HHD (i.e. scanning) was avoided.

As we described in section 4.4, because of sensor limitations or possible malfunctions, not all cues can be precise. We expected that less precise cues would increase detection latency, but we wanted to determine if we could forestall the costs by signaling the imprecision (dashed versus solid cueing), possibly broadening the attention field. Our attempts at broadening attention were not successful in a way that might have been predicted by the zoom lens model (Eriksen and Yeh, 1985; Eriksen and St. James, 1986), because it appears that the cueing imprecision (dashed symbology) does not change where a serial search pattern starts (at the center). If presenting the cue dashed would have had such a broadening effect as expected, we should have observed slower detection times in the two accurate trials with dashed cues (i.e. spotlight breadth costs), and we should have been discovered a reduction in type A costs (discussed below). Our results show neither an increase in detection latency nor a reduction in the type A costs for the dashed cue on the trails when it was accurate. Cue imprecision however,

led to progressively greater cost in the time of detection of less salient targets (landmines), which being less visible in the periphery (smaller UFOV), suffered more as this peripheral area was widened. While this cost for imprecise cueing was harmful for the low salience targets, the imprecise cueing remained beneficial for quicker detection times of the more salient target type, even when the cue was off by as much as 45 degrees from the center of the target.

The type A costs had been reported by Yeh et al., (in press) as characterized by the decreased detection of high priority but unexpected targets presented concurrently with an expected target and particularly when the latter was cued. This type A cost of cueing was replicated here. Unlike Yeh, Wickens, and Seagull (in press) who found that cueing enhanced the type A costs only with the HMD and not the HHD, we found here that the type A costs were roughly equivalent in their magnitude across platforms (only slightly greater with HMD and then only when the cue was solid). We attribute this muting of platform differences to our more careful equating the visibility of the HHD and the HMD, and equalizing the weight of the subject's heads for both platforms (therefore making search strategies more similar). That is, given that our HHD was more visible (than in Yeh et al.), subjects were somewhat more likely to rely upon its cue, and hence suffer the type A costs of that reliance. Thus our effects while not identical to Yeh, Wickens, and Seagull (in press) are reconcilable through simple procedural differences.

We also anticipated that these type A costs would be manifest in other aspects of data, particularly in the signaling of imprecision through the dashed cue and in the destruction of trust. However as noted, the dashed cues, a display feature, did not appear to broaden attention (hence, attenuate type A costs in this regard). This failure of the display feature to attenuate attention, replicates Gempler and Wickens (1998) who also found that some techniques of displaying

uncertainty (such as the wedge shaped 95% confidence interval used in their CDTI display) are not always successful and is also consistent with the search pattern described above, where by scanning starts at the center, independently of solid or dashed cue presentation. However, there was some evidence that type A costs (and therefore attention breadth) was modulated by experience of catastrophic failures, the issue to which we now turn.

When subjects' trust of the automation was betrayed (trial 4 of the last block), we observed extremely long detection times because of the effects of overtrust in the system automation. Not surprisingly, the initial system failure was met quite unexpectedly by subjects. The long search times in the cued regions (soemtimes greater than a minute) would be expected from users who have built trust in a system over time (Parasuraman and Riley, 1997; Lee and Moray, 1991; Lee and Moray, 1992; Muir and Moray, 1996; Singh, et al., 1993). Interestingly, longer detection times were recorded for the HHD than for the HMD, a fact we can attribute to the less accurate precision of valid cueing in HHD because of vertical ambiguity (no y-axis information); hence it took longer to notice the failure while using an HHD. In the HMD condition, the absence of a target within or right next to the solid cueing symbology should have immediately raised some level of suspicion. However, two of the subjects' suspicion regarding cue validity was never raised, as they saw 'ghost targets'. This detection of a target that is not present is a clear example of the type T cost. A non-target was cued and the subject's high level of expectancy led him to see a target or infer an event that was not there, similar to findings by Conejo and Wickens (1997) and Mosier et al. (1998). This seemingly high level of overtrust in the automation, coupled with a high level of expectancy, occurred on only two occasions, both with subjects who were using the HMD platform.

Erroneous cueing that was imposed two trials after the first erroneous cue trial (trial 7) did not produce nearly as lengthy search times in the incorrect location. Subjects had been 'burned once' and presumably abandoned the incorrect search space more quickly after a brief confirmation of the cue's erroneous nature, this confirmation again occurring more rapidly in the HMD condition. After this second erroneous cue presentation, the user's trust in the automation seemed to effect their search strategy even more as they take longer than in previous blocks (before automation failure) to find accurately cued targets. Subjects always initially followed the cue even after erroneous trials, which was usually a good strategy, since even during block 11, there were more accurately cued trials than erroneous trials and without the cue subjects had no other evidence to direct their initial search direction, (i.e., using Fisher's characterization of highlighting validity, we might say the cue validity was approximately 70%.) Subjects attained a somewhat "calibrated trust" by the end of the experiment as detection times moved more closely to the average detection times obtained before automation failure, allowing them to realize some, but not maximum benefits of cueing.

We anticipated that the catastrophic loss of trust would expand the attention (search) field, and that this might mitigate type A costs. This mitigation was indeed observed. Destroying the users' trust in the automation seemed to greatly effect the way that the subjects modulated their visual search strategies or possibly even their attention breadth. The type A cost associated with attentional cueing decreased significantly following the trials where trust was 'betrayed' (see section 7.7). The higher detection rate of the unexpected event (91% after, versus 50% before) implies that subjects employed a different strategy when there was a lower level of trust in the automated cueing system. This strategy resulted in them being less immediately

drawn to the cue, less likely to report a target there, and more likely to search carefully elsewhere for the higher priority nuclear device.

In conclusion, our results are encouraging with regards to attentional cueing. While the highest levels of cue imprecision did not benefit performance in the detection task relative to the uncued condition for targets with low salience (landmines), cues containing imprecision still proved beneficial in the detection of the more salient targets. Potentially systems that can reduce search space within a diameter of 90 degrees or less could provide benefits for target detection, and these benefits can be realized with either type of display platform. What is not known is whether our explicit display of cue precision extended these limits of cue imprecision, since we had no trials that presented a solid cue with imprecision (until our catastrophic failure). This is an area that needs further research.

Although HMDs provided faster detection times for cued targets, clutter produced by the HMD symbology in the FFOV proved to be costly for detecting low acuity targets in the uncued trials. If more uncued targets are suspected than cued ones, it might prove beneficial to temporarily reduce some of the information present in the FFOV. However, the ego-referenced symbology presented in the FFOV proved beneficial for azimuth judgement and reduced scanning. Careful consideration of these performance tradeoffs should be considered before committing to either display platform.

Trust in the automation greatly effects the way in which users will interact with the system, with our findings suggesting that attention breadth and/or search strategy is somewhat modulated by user trust in the system. The long term effects of unreliability did not seem to be as detrimental or as long lasting as we anticipated, that is by the end of block 11 we had abolished type A costs but retained some (albeit reduced) cue benefits, as subject's trust became

more accurately calibrated to the system. A high catastrophic failure rate in the automation may prevent users from ever reaching the maximum benefits of an automated system (undertrust), whereas the costs of overtrust can be even more costly, with the greater possibilities of type A and T costs. These costs, A and T, should be carefully considered in the implementation of any type of automated cueing system, as well, the magnitude of these costs should be taught in training, possibly inducing errors in the automation during training to assist the user in properly calibrating his or her trust.

References

Andre, A.D. and Cashion, P.A. (1993). Visual scanning in the functional visual field. Society for Information Display International Symposium Digest of Technical Papers (pp. 114-117). Playa Del Rey, CA: SID.

Arditi, A. (1986). Binocular vision. In K.R. Boff, L. Kaufman, and J.P. Thomas (Eds.), Handbook of Perception and Human Performance. Vol 1, New York: John Wiley.

Becklen, R. and Cervone, D. (1984). Selective looking and the noticing of unexpected events. *Memory and Cognition*. **11**(6), 601-608.

Behringer, R. (1999). A hybrid system for outdoor augmented reality. *Proceedings of the Third Annual Laboratory Symposium on Advance Displays and Interactive Displays*. University Park, Maryland.

Brown, T.J. (1991). Visual display highlighting and information extraction. *Proceedings of the Human Factors Society 35th Annual Meeting.* Santa Monica, CA: Human Factors Society.

Converse, S.A., Kozai, S., & Batten, D. (1992). Color coding to facilitate performance of focused attention tasks with object displays. *Proceedings of the Human Factors Society 36th Annual Meeting.* Santa Monica, CA: Human Factors Society.

Crowder, R.G. (1993). Short-term memory: Where do we stand?. *Memory and Cognition*, **21**(2), 142-145.

Donner, K.A., McKay, T., O;Brien, K.M., & Rudisill, M. (1991). Display format and highlighting validity effects on search performance using complex visual displays. . *Proceedings of the Human Factors Society 35th Annual Meeting.* Santa Monica, CA. Human Factors Society.

Drury, C.G., and Clement, M.R. (1978). The effect of area, density, and number of background characters on visual search. *Human Factors*, **20**, 597-602.

Duncan, J. (1984). Selective attention and the organization of vital information. *Journal of Experimental Psychology: General.* **113**, 501-517.

Edworthy, J., Loxley, S., & Dennis, I. (1991). Improving auditory warning design: relationships between warning sound parameters and perceived urgency. *Human Factors*, **33**(2), 205-231.

Endsley, M.R. (1995). Measurement of situation awareness in dynamic systems. *Human Factors*, **37**(1), 65-84.

Eriksen, C.W., and Webb, J.M. (1989). Shifting of attentional focus within and about a visual display. *Perception and Psychophysics*, **45**(2), 175-183.

Fischer, E., Haines, R.F., and Price, T.A. (1980). Cognitive issues in head-up displays (NASA Technical Paper 1711). Moffet Field, CA: NASA Ames Research Center.

Fisher, D.L., Coury, B.G., Tengs, T.O., & Duffy, S.A. (1989). Minimizing the time to search visual displays: The role of highlighting. *Human Factors*, **31**(2), 167-182.

Fisher, D.L. and Tan, K.C. (1989) Visual displays: The highlighting paradox. *Human Factors*, 31, 17-30.

Hofer, E.F., Palen, L.A., and Possolo, A. (1993). Flight deck information management: An experimental study of functional integration of approach data. *Proceedings of the 7th International Symposium on Aviation Psychology*. Columbus, OH: Department of Aviation, Ohio State University.

Kantowitz, B.H., Hanowski, R.J., & Kantowitz, S.C. (1997). Driver acceptance of unreliable traffic information in familiar and unfamiliar settings. *Human Factors*. 39(2), 164-176.

Kirschenbaum, S.S., & Arruda, J.E. (1994). Effects of graphic and verbal probability information on command decision making. *Human Factors*. 36(3), 406-418.

Kramer, A.F. and Jacobson, A. (1991) Perceptual organization and focused attention: The role of objects and proximity in visual processing. *Perception and Psychophysics*, **50**, 267-284.

Larish, I. and Wickens, C.D. (1991). Divided attention with superimposed and separated imagery: Implications for head-up displays. University of Illinois Institute of Aviation Technical Report (ARL-91-4/NASA-HUD-91-1). Savoy, IL: Aviation Research Laboratory.

Lee, J. & Moray, N. (1992). Trust, control strategies and allocation of function in human-machine systems. *Ergonomics*. **35**(10), 1243-1270.

Lee J. & Moray, N. (1994). Trust, self confidence, and operators' adaptation to automation. *Int. J. Human-Computer Studies*. 40, 153-184.

Madden, D.J., and Allen, P.A. (1989). Amount and duration of attentional demands during visual search. *Perception and Psychophysics*, **45**(6), 577-585.

May, P. and Wickens, C.D. (1995). The role of visual attention in head-up display: Design implications for varying symbology intensity. University of Illinois Institute of Aviation Technical Report (ARL-95-3/NASA-HUD-95-1). Savoy, IL: Aviation Research Laboratory.

Mosier, K.L., Skitka, L.J., Heers, S., and Burdick, M. (1998). Automation bias: Decision making and performance in high-tech cockpits. *The International Journal of Aviation Psychology* 8(1):47-63.

Muir, B.M., and Moray, N. (1996). Trust in Automation. Part II. Experimental studies of trust and human intervention in a process control simulation. *Ergonomics.* 39(3), 429-460.

Muller, H.J. and Rabbitt, P.M. (1989). Spatial cueing and the relation between the accuracy of "where" and "what" decisions in visual search. *Quarterly Journal of Experimental Psychology* Vol 41A(4) 747-773.

Parasuraman, R., and Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors*, **39**(2) 230-253.

Posner, M.I. (1978). *Chronometric explorations of mind.* Hillsdale, N.J. Erbaum.

Posner, M.I., Snyder, C.R.R., and Davidson, B.J. (1980). Attention and the detection of signals. *Journal of Experimental Psychology: General.* 109(2), 160-174.

Rizzolatti, G., Riggio, L., Dascola, I., and Umiltá, C. (1987). Reorienting attention across the horizontal and vertical meridians: Evidence in favor of a premotor theory of attention. *Neuropsychologia*, 25, 31-40.

Rossoe, S.N. (1987a) The trouble with HUDs and HMDs. *Human Factors Society Bulletin*, 30(7), 1-3.

Sampson, J. B. (1993) Cognitive Performance by Individuals Using a Head-Mounted-Display While Walking. *In Proceedings of the Human Factors and Ergonomics Society 37th Annual Meeting.* Santa Monica, CA: Human Factors and Ergonomics Society.

Scanlan, L.A. (1977). Target acquisition in realistic terrain. *Proceedings of the Human Factors Society 21st Annual Meeting.* Santa Monica, CA: Human Factors Society.

Singh, I.L., Malloy, R. and Parasuraman, R. (1993). Individual differences in monitoring failures of automation. *The Journal of General Psychology* 120(3), 357-373.

Smith, S.L. and Goodwin, N.C. (1971). Blink coding for information displays. *Human Factors*, 14, 345-347.

Sorkin, R.D., Kantowitz, B.H., Kantowitz, S.C. (1988). Likelihood alarm displays. *Human Factors*, **30**(4), 445-459.

Tan, K.C. and Fisher, D.L. (1987). Highlighting and search strategy considerations in computer generated displays. *Proceedings of the Human Factors Society 31st Annual Meeting*. Santa Monica, CA: Human Factors Society.

Tversky, A. and Kahneman, D. (1984). Judgement under Uncertainty: Heuristics and Biases. *Science* vol 185, 1124-1131.

Ververs, P.M. and Wickens, C.D. (1998). Head-up displays: Effects of clutter, display intensity, and display location on pilot performance. *The International Journal of Aviation Psychology*, **8**(4), 377-403.

Wickens, C. D. (1992). *Engineering Psychology and Human Performance* (2nd ed.). New York: HarperCollins.

Wickens, C.D., (1980) The structure of attentional resources. In R. Nickerson (ed.), *Attention and Performance VIII* (pp. 239-257). Hillsdale, NJ: Erlbaum.

Wickens, C.D. (1984) Processing resources in attention. In R. Parasuraman & R. Davies (eds.), *Varieties of attention* (pp.63-101). New York: Academic Press.

Wickens, C.D., and Carswell, C.M. (1997). Information processing. In G. Salvendy (ed.), *Handbook of Human Factors & Ergonomics*. (2nd ed), New York: John Wiley. 1997.

Wickens, C.D., Gordon, S.E., & Liu, Y. (1998). *An Introduction to Human Factors Engineering*. New York: Wesley Longman, Inc.

Wickens, C.D. and Long J. (1995) Object versus space-based models of visual attention: Implications for the design of head-up displays, *Journal of Experimental Psychology: Applied*, **1**, 3, 1979-183.

Wickens, C.D., & Morphew, E. (1997). *Predictive features of a cockpit traffic display: A workload assessment.* University of Illinois Institute of Aviation Technical Report (ARL-97-6/NASA-97-3). Savoy, IL: Aviation Res. Lab.

Wickens, C.D., Pringle, H. & Merlo, J. (1999). Integration of Information Sources of Varying Weights: The Effect of Display Features and Attention Cueing (Technical Report ARL-99-1/ARMY-FED-LAB-99-1). Savoy, IL: University of Illinois, Institute of Aviation, Aviation Research Laboratory.

Yantis, S. (1992). Multielement visual tracking: Attention and perceptual organization. *Cognitive Psychology*. **24**, 295-340.

Yeh, M. & Wickens, C.D. (1997). Effects of Color Coding, Intensity Coding, and Decluttering on Visual Search in Electronic Map Displays (Technical report) Savoy, IL: University of Illinois, Institute of Aviation, Aviation Research Laboratory..

Yeh, M., Wickens, C.D., & Seagull, F.J. (1998). *Effects of frame of reference and viewing condition on attentional issues with helmet-mounted displays* (Technical Report ARL-98-1/ARMY-FED-LAB-98-1). Savoy, IL: University of Illinois, Institute of Aviation, Aviation Research Laboratory.

Yeh, M. & Wickens, C.D. (1998). Visual Search and Target Cueing: A comparison of head-mounted versus hand-held displays on the allocation of visual attention (Technical Report ARL-98-2/ARMY-FED-LAB-98-2). Savoy, IL: University of Illinois, Institute of Aviation, Aviation Research Laboratory.

Yeh, M., Wickens, C.D., & Seagull, F.J. (in press). Target cueing in visual search: The effects of conformality and display location on the allocation of visual attention. *Human Factors.*